Ionospheric modelling and Data Assimilation

Da Dalt, Federico

Award date:
2015

Awarding institution:
University of Bath

IONOSPHERIC MODELLING

AND DATA ASSIMILATION

Federico Da Dalt

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Electronic and Electrical Engineering

April 2015

Acknowledgments

I am deeply grateful to my supervisor, Prof. Cathryn Mitchell, whose expertise, patience and generous guidance have been invaluable to the success of my research. Cathryn has provided me with incredible opportunities throughout my PhD that have greatly developed me both personally and professionally.

I would like to thank the members of the Invert group of the University of Bath and in particular Dr. Nathan Smith, Dr. Mark Greco and Tommaso Panicciari for their precious assistance.

I must express my gratitude to international scientists such as Prof. Jan J. Sojka (Centre for Atmospheric and Space Sciences, Utah State University), Dr. Anthea Coster and Dr. Phil Erickson (MIT Haystack Observatory) for their constructive comments and suggestions.

Global Positioning System (GPS) data were provided by the Crustal dynamics data information system (NASA), the International GNSS Service (IGS), the Ecole Nationale des Sciences Géographiques IGS service, the University Navstar Consortium (UNAVCO), and the Garner GPS archive. Radar observations and analysis at Millstone Hill are supported under Cooperative Agreement with the Massachusetts Institute of Technology (MIT) by the United States (US) National Science Foundation. Ionosonde observations are taken from the Space Physics Interactive Data Resource (SPIDR). Space weather indices were provided by the National Geophysical Data Center (NGDC) of the National Oceanic and Atmospheric Administration (NOAA)

This research activity was supported by a Marie Curie initial training network (TRANSMIT) within the 7th European Community Framework Programme.

Finally, I thank my family for its continuous support and encouragement.

Abstract

A New Ionospheric Model (ANIMo) based upon the physics of production, loss, and vertical transport has been developed. The model is driven by estimates of neutral composition, temperature and solar flux and is applicable to the mid-latitude regions of the Earth under quiet and moderate geomagnetic conditions.

This model was designed to exhibit specific features that were not easy to find all together in other existing ionospheric models. ANIMo needed to be simple to use and interact with, relatively accurate, reliable, robust and computationally efficient. The definition of these characteristics was mostly driven by the intention to use ANIMo in a Data Assimilation (DA) scheme. DA or data ingestion can be described as a technique where observations and model realizations, called background information, are combined together to achieve a level of accuracy that is higher than the accuracy of the two elements taken separately. In this project ANIMo was developed to provide a robust and reliable background contribution. The observations are given by the Global Positioning System (GPS) ionospheric measurements, collected from several networks of GPS ground-station receivers and are available on on-line repositories.

The research benefits from the Multi-Instrument Data Analysis System (MIDAS) [*Mitchell and Spencer*, 2003; *Spencer and Mitchell*, 2007], which is an established ionospheric tomography software package that produces three dimensional reconstructions of the ionosphere starting from GPS measurements. Utilizing ANIMo in support of MIDAS has therefore the potential to generate a very stable set-up for monitoring and study the ionosphere. In particular, the model is expected to compensate some of the typical limitations of ionospheric tomography techniques described by *Yeh and Raymund* [1991] and *Raymund et al.* [1994]. These are associated with the lack of data due to the uneven distribution of ground-based receivers and limitations to viewing angles.

Even in regions of good receiver coverage there is a need to compensate for information on the vertical profile of ionisation. MIDAS and other tomography techniques introduce regularization factors that can assure the achievement of a unique solution in the inversion operation. These issues could be solved by aiding the operation with external information provided by a physical model, like ANIMo, through a data ingestion scheme; this ensures that the contribution is completely independent and there is an effective accuracy improvement. Previously, the limitation in vertical resolution has been solved by applying vertical orthonormal functions based upon empirical models in different ways [*Fougere*, 1995; *Fremouw et al.*, 1992; *Sutton and Na*, 1994]. The

potential for the application of a physical model, such ANIMo is that it can provide this information according to the current ionospheric conditions.

During the project period ANIMo has been developed and incorporated with MIDAS. The result is A New Ionospheric Data Assimilation System (ANIDAS); its name suggests that the system is the implementation of ANIMo in MIDAS. Because ANIDAS is a data ingestion scheme, it has the potential to be used to perform not only more accurate now-casting but also forecasting. The outcomes of ANIDAS at the current time can be used to initialise ANIMo for the next time step and therefore trigger another assimilation turn. In future, it is intended that ANIMo will form the basis to a new system to predict the electron density of the ionosphere – ionospheric forecasting.

Table of contents

List of tables

List of figures

List of abbreviations

(In alphabetic order)

3DVar	Three-Dimensional Variational Data Assimilation
4DVar	Four-Dimensional Variational Data Assimilation
AFRL	Air Force Research Laboratory
AFWA	Air Force Weather Agency
AIAA	American Institute of Aeronautics and Astronautics
ANIDAS	A New Ionospheric Data Assimilation System
ANIMo	A New Ionospheric Model
ANSI	American National Standard Institute
BLUE	Best Linear Unbiased Estimator
C/A	Coarse Acquisition code
CCMC	Community Coordinated Modelling Center
CIRA	COSPAR International Reference Atmosphere
CIRA-86	1986 version of the CIRA model
CIT	Computerized Ionospheric Tomography
CITFM	Coupled Ionosphere Thermosphere Forecast Mode
CMAT	Coupled Middle Atmosphere and Thermosphere
CMAT2	Coupled Middle Atmosphere and Thermosphere-2
CME	Coronal Mass Ejection
CMIT	Coupled Magnetosphere-Ionosphere-Thermosphere
COSMIC	Constellation Observing System for Meteorology
COSPAR	Committee on Space Research

CTIM	Coupled Thermosphere-Ionosphere Model
CTIP	Coupled Thermosphere-Ionosphere-Plasmasphere
CTIPe	Coupled Thermosphere-Ionosphere-Plasmasphere Electrodynamics
DA	Data Assimilation
DMSP	Defense Meteorological Satellite Program
DoD	Department of Defense
DSV	Singular Value Decomposition
EDAM	Electron Density Assimilative Model
EOF	Empirical Orthonormal Function
EUV	Extreme Ultra Violet
EUVAC	A solar EUV Flux Model for aeronomic calculations
FLIP	Field Line Interhemispheric Plasma
FTCS	Forward in Time Centred in Space
GAIA	Ground-to-topside model of Atmosphere and Ionosphere for Aeronomy
GAIM	Global Assimilation of Ionospheric Measurements
GAIM	Global Assimilative Ionospheric Model
GIRO	Global Ionospheric Radio Observatory
GNSS	Global Navigation Satellite System
GPID	Global Plasmasphere Ionosphere Density
GPS	Global Positioning System
GTIM	Global Theoretical Ionospheric Model
GTIM	Global Ionosphere Thermosphere Model
HWM	Horizontal Wind Model
HWM07	2007 version of HWM

IAGA	International Association of Geomagnetism and Aeronomy
ICTP	International Centre for Theoretical Physics
IDA3D	Ionospheric Data Assimilation Three-Dimensional
IDA4D	Ionospheric Data Assimilation Four-Dimensional
IFM	Ionospheric Forecast Model
IGRF	International Geomagnetic Reference Field
IGS	International GNSS Service
IMF	Interplanetary Magnetic Field
IMM	Ionosphere-Magnetosphere Model
IPM	Ionosphere-Plasmasphere Model
IRI	International Reference Ionosphere model
IRI-2012	2012 version of the IRI model
IRI-95	1995 version of the IRI model
ISR	Incoherent Scatter Radar
JPL	Jet Propulsion Laboratory
LMT	Local Magnetic Time
LT	Local Time
MIDAS	Multi-Instrument Data Analysis System
MIT	Massachusetts Institute of Technology
MSIS	Mass Spectrometer Incoherent Scatter
MSIS-86	1986 version of the MSIS model
MSTU	Moscow State Technical University
MTIE-GCM	Magnetosphere-Thermosphere-Ionosphere Electrodynamics General Circulation Model
MURI	Multidisciplinary University Research Initiatives

NASA	National Aeronautics and Space Administration
NAVSTAR	Navigation Satellite Timing and Ranging System
NCAR	National Center for Atmospheric Research
NICT	National Institute of Information and Communications Technology
NOAA	National Oceanic and Atmospheric Administration
NRL	Naval Research Laboratory
NRLMSISE-00	2000 version of the MSIS model
NWF	Numerical Weather Forecasting
P or Y	Precision code
PIM	Parameterized Ionospheric Model
PRISM	Parameterized Real-Time Ionospheric Specification Model
RIM	Ridley Ionosphere Model
RINEX	Receiver Independent Exchange Format
RMS	Root Mean Square
SAMI2	Sami2 is Another Model of the Ionosphere
SAMI3	Sami3 is Another Model of the Ionosphere
SLIM	Semi-Empirical Low-Latitude Ionospheric Model
SPIDR	Space Physics Interactive Data Resources
SSULI	Special Sensor Ultraviolet Limb Imager
SSUSI	Special Sensor Ultraviolet Spectrographic Imager
STEC	Slant Total Electron Content
SUPIM	Sheffield University Plasmasphere-Ionosphere Model
T/ICT4D	Telecommunication and Information & Communication Technology for Development
TDIM	Time Dependent Ionospheric Model

TEC or TEC	Total Electron Content
TECU	Total Electron Content Unit
TFM	Thermosphere Forecast Model
TGCM	Thermosphere General Circulation Model
TIE-GCM	Thermosphere-Ionosphere-Electrodynamics General Circulation Model
TIGCM	Thermosphere-Ionosphere General Circulation Model
TIME-GCM	Thermosphere-Ionosphere-Mesosphere-Electrodynamics General Circulation Model
TING	Thermosphere-Ionosphere Nested Grid
TRANSMIT	Training Research and Applications Network to Support the Mitigation of Ionospheric Threats
UAM	Upper Atmosphere Model
UCAR	University Corporation for Atmospheric Research
UCL	University College of London
URSI	International Union of Radio Science
US	United States
USC	University of Southern California
USU	Utah State University
UV	Ultra Violet
VTEC	Vertical Total Electron Content

Notations and symbols

(In order of appearance)

K	Kennziffer
K_p	Planetarische Kennziffer
D_{st}	Disturbance Storm Index
$F10.7$	10.7 cm solar flux
a_p	3-hourly K_p equivalent planetary index
A_p	Daily K_p equivalent planetary index
B_z	Vertical component of the Interplanetary Magnetic Field (IMF)
$*$	Excited state
$e^- p$	Solar or auroral electron (primary)
$e^- s$	Electron released from neutral species (secondary)
h	Planck's constant
ν	Frequency of the electromagnetic wave of an absorbed photon
N_e or N_e	Electron density
t	Time
Q	Production term
L	Loss term
v	Generic velocity
$\vec{B}$	Magnetic field
$\vec{E}$	Electric field
ExB	Electromagnetic drift

v_{drift}	Electromagnetic drift velocity
$v_{diffusion}$	Diffusion velocity
D_a	Ambipolar diffusion coefficient
z	Altitude
I	Geomagnetic dip angle
m	Mass
g	Gravity's acceleration
k_B	Boltzmann's constant
T	Temperature
T_i or T_i	Ion temperature
T_e or T_e	Electron temperature
m_i	Mass of a generic ion
v_{in}	Ion-neutral collision frequency
$v_{neutral\ wind}$	Neutral wind velocity
V_{wind}	Wind speed
k_L	Constant for the Loss term
P	Pseudo-range
Φ	Carrier phase
ρ	Real distance (range)
c	Speed of light
dt	Receiver clock offset
dT	Satellite clock offset
d_I	Ionospheric error

d_T	Tropospheric error
d_M	Multipath effect error
h_s	Satellite hardware dispersive component
h_r	Receiver hardware dispersive component
λ_s	Wavelength of the signal
N	Carrier phase ambiguity
ϵ	Residual error
RX	Receiver position
TX	Transmitter position
s	Satellite ray path
n^2	Refractive index
X, Y, Z	Magnetoionic paramters
e	Electron charge
ε_0	Permittivity of the free space
m_e	Electron mass
ω	Angular wave frequency
k	Constant value
f	Signal frequency
P_{L2}	Pseudo-range for L2 GPS frequency signal
P_{L1}	Pseudo-range for L1 GPS frequency signal
I_p	Ionospheric parameter
f_{L2}	L2 GPS frequency
f_{L1}	L1 GPS frequency

b_r	Bias from the hardware dispersive components of the receiver
b_s	Bias from the hardware dispersive components of the satellite
Φ_{L1}	Carrier phase for L1 GPS frequency signal
Φ_{L2}	Carrier phase for L2 GPS frequency signal
λ_{L1}	Wavelength for L1 GPS frequency signal
N_{L1}	Phase ambiguity for L1 GPS frequency signal
λ_{L2}	Wavelength for L2 GPS frequency signal
N_{L2}	Phase ambiguity for L2 GPS frequency signal
h'	Ionosonde virtual height of reflection
h_m	Ionosonde real height of reflection
f_oF2	Critical frequency of the F2 layer
f_o	Critical frequency for ordinary propagation
f_x	Critical frequency for extraordinary propagation
N_m	Electron density at the peak
h_mF2 or hmF2	Electron density peak altitude for the F2 layer
$N_mF2$or NmF2	Electron density at the peak for the F2 layer
σ_e	Radar cross section associated with an electron (Thomson's cross section)
r_e	Effective radius of one electron
v_T	Thermal speed
Δf_e	Doppler shift of an electron moving toward the radar
λ_R	Wavelength of the radar signal
λ_D	Debye length
Λ	Wavelength of a specific acoustic wave (where $\Lambda = \frac{1}{2}\ \lambda_R$)

F	Frequency of an acoustic wave
F^+	Frequency of an ion-acoustic wave
F^-	Frequency of an electron-acoustic wave
f_p	Plasma frequency
N_i	Ion density
N_n	Neutral density
T_n	Neutral temperature
k_R	Recombination rate coefficient
Wd	Vertical diffusion velocity implemented in ANIMo
T_p	Plasma temperature
m_{O^+}	Monoatomic oxygen ion mass
ν_{O^+-O}	Ion-neutral atomic oxygen collision frequency
T_r	Temperature parameter for the collision frequency calculation
Q_{O^+}	Production term for the monoatomic oxygen ion
L_{O^+}	Loss term for the monoatomic oxygen ion
i	Spatial (cell) index (generic dimension)
Δt	Time interval
Δz	Space interval
J	Cost function
$\boldsymbol{x}$	State vector
$\boldsymbol{y}$	Observations vector
$\mathbf{H}$	Observation operator
$\boldsymbol{\Sigma}$	Observation covariance matrices
$\boldsymbol{x}_b$	Background or *a priori* state

$\mathbf{B}$	Background covariance matrices
J_M	MIDAS cost function
$\boldsymbol{z}$	Ionospheric observation vector taken by GPS ground based receivers
λ	MIDAS regularization tuning parameter
$\mathbf{K}$	Vertical basis function operator
$\boldsymbol{n}$	Electron density vector
$\delta\boldsymbol{x}$ or $\delta\boldsymbol{x}_r$	Residual state vector
$\delta\widehat{\boldsymbol{x}}$ or $\delta\widehat{\boldsymbol{x}}_r$	Residual pseudo-solution or analysis increment
$\boldsymbol{n}_b$	Background state produced by ANIMo
$\widehat{\boldsymbol{n}}$ or $\widehat{\boldsymbol{n}}_r$	Analysis vector of ANIDAS
$\boldsymbol{z}_b$	Ionospheric observation vector modelled by ANIMo
$\delta\boldsymbol{z}$	Observation residual
J_A or $J_{A\text{-}r}$	ANIDAS cost function
α	ANIDAS background covariance matrix tuning parameter
$\widehat{\boldsymbol{x}}_{nr}$	Non-residual pseudo-solution
$J_{A\text{-}nr}$	Cost function of ANIDAS non-residual version
$\boldsymbol{x}_{nr}$	Non-residual state vector
$\widehat{\boldsymbol{n}}_{nr}$	Analysis vector of ANIDAS non-residual version
$\boldsymbol{x}_{\mathrm{t}}$	True state
ε_{b}	Background error
e_i	Scaled background error
ρ	Correlation
j	Generic dimension

$\mathbf{T}$	Toeplix matrix
t_{cc}	Time correlation coefficient
σ^2	Variance
$J_{M\text{-}r}$	Cost function of MIDAS residual version
$J_{M\text{-}nr}$	Cost function of MIDAS non-residual version

Chapter 1 Introduction

The first transatlantic message was broadcasted using radio waves by Guglielmo Marconi in the 1901. The broadcast was made possible because of the presence of the ionosphere, the conducting layer of the atmosphere whose existence was officially confirmed a few decades later by Edward V. Appleton. This layer, as its name suggests, is populated by both neutral and charged particles and lies between 70 and 1000 km from the Earth's surface. The ionosphere plays an important role on radio communications; therefore it has been extensively studied and investigated since its discovery.

Nowadays the ionosphere is monitored and studied especially because of its degrading effects on satellite communications. For example, in the positioning applications of the Global Navigation Satellite Systems (GNSS) the ionosphere could represent the larger source of error if uncorrected. Travelling through the ionosphere causes time and phase delays on satellite signals. The degree of these delays is directly proportional to the concentration of electrons that lie in the satellite-receiver signal path; therefore it is possible to estimate ionospheric measurements from GNSS receivers.

Computerized Ionospheric Tomography (CIT) is a technique that collects and, by means of an operation called tomographic inversion, combines ionospheric measurements to compute three dimensional electron density reconstructions of the ionosphere. The limitations of this technique are related to the lack of information due to the sparse distribution of ground receivers and the limited angle of their measurements.

The established CIT software package MIDAS deals with these issues by applying regularization methods and incorporating a priori information in the inversion. Thus far, this information has been provided by empirical models.

This dissertation aims to present a novel approach for supporting CIT that consists in the creation of a physics-based ionospheric model and its implementation in a Data Assimilation approach. The latter has the potential to not only improve the imaging but also to produce short term forecasting of the ionosphere. In order to reach these over-arching goals, the project tasks were organized to achieve intermediate objectives.

The intermediate objectives of the presented project can be summarised as follows:

- Realization of a physics based ionospheric model: The model needs to exhibit specific requirements. First of all it must not be too complicated in its design. This means a small

number of input parameters and a relatively easy usage format without compromising the accuracy of the model output. Secondly, the model needs to be robust and reliable as it can be used to simulate unsettled conditions of the ionosphere. Certain efficiency in terms of computing effort is also required.

- Assessment of the model: This consists in understanding if the model is suitable for imaging and forecasting purposes. Validation tests are necessary to evaluate the accuracy of the model against measured and simulated data. Complementary tests are needed to assess the aforementioned requirements.
- Realization and implementation of an ionospheric DA scheme: Where the model is supposed to produce the background state. The observations are GPS measurements of the ionosphere taken from ground based receivers. The output, known as the analysis, is expected to be a more accurate image of the ionosphere with respect to an output produce solely by the model or extrapolated solely from the observation. The approach could be used in an iterative manner to provide combined reconstructions for multiple consecutive times (now-casting mode). In addition, the scheme can be used to provide a better initialization for the model to produce predictions of the ionosphere (forecasting mode).
- Validation of the ionospheric DA scheme: This involves a comparison between real data and the DA outputs applied in both now-casting and foresting modes.

This dissertation is composed of nine chapters and one appendix; including this introduction. Excluding this introduction and the conclusions chapters, the structure of this dissertation can be broadly divided into two parts. The first part refers to the background and the second part presents novel methodologies and their results. In particular, Chapter 2 briefly describes the Sun-Earth relationships and the main features of the ionosphere by focussing on processes and driving forces that have been relevant for the development of A New Ionospheric Model (ANIMo). An explanation of the different data sources used in this project is given in Chapter 3; this includes the physics basis of some of the instruments that provided the adopted measurements. Chapter 4 is a literature review of the existing ionospheric models and a few noteworthy DA approaches. The models are listed in a chronological order that considers possible evolutions and upgrades of the models while the DA schemes are treated separately. Chapter 5 introduces ANIMo, describes its characteristics and thoroughly explains its functioning. Validation tests were conducted by comparing ANIMo outputs against ionospheric measurements and modelled simulations. Chapter 6 shows the results of these tests for five case studies set in the same location. Four of them were selected in similar low geomagnetic activity conditions, but at four different periods of the year (one for each season). The fifth case refers to the winter season in perturbed ionospheric conditions. Sensitivity tests of ANIMo forcing parameters are also included for the winter case study. Chapter 7 is dedicated to the novel ionospheric DA approach called A New Ionospheric Data Assimilation Scheme (ANIDAS). This technique and its important elements, such as the

generation of covariance matrixes, are described. The results are shown and discussed in Chapter 8. The winter case study is once again selected in the validation test which involves the usage of ANIDAS in now-casting and forecasting modalities. The final chapter collates the conclusions of the dissertation and incorporates future developments of the project.

Chapter 2 The Earth's Ionosphere

Abstract

The terrestrial ionosphere is conventionally defined as the portion of the upper atmosphere that is ionised by solar and cosmic radiation. It lies between approximatively 70 and 1000 km and it is characterised by the presence of different ionospheric regions called layers. The structure of the ionosphere is highly variable and it depends on many factors. The principal processes of its formation are related to the ionizing action of solar radiation and the chemical reactions that determine the duration of ion and electrons. These charged particles are very sensitive to transportation processes, heating, and geomagnetic phenomena.

Introduction

This project required extensive knowledge of the ionospheric medium. Three aspects were particularly relevant. The first is related to understanding the principal mechanisms of ionospheric dynamics (such as its formation, maintenance etc.), which has been fundamental in order to build the ionospheric model ANIMo. The second aspect is also related to the modelling. The simulations of the mechanisms described are controlled by forcing parameters. These are mostly measurements, or indices extrapolated from measurements, of real phenomena of the Sun-Earth system. The ability to use them within a model requires a good understanding of the phenomenon that they describe. The third aspect is much broader. A sound knowledge of ionospheric features, especially the ionosphere's structure, was essential to assess the validity of the proposed methods.

This chapter briefly introduces the essential background about the ionosphere. Section 2.1 refers to the Sun-Earth system, starting from the Sun (Subsection 2.1.1), continuing with the concepts of solar emissions and Interplanetary Magnetic Field (IMF) (Subsection 2.1.2) and ending at the geomagnetic field (Subsection 2.1.3). Section 2.2 introduces the concept of ionosphere while Section 2.3 describes the ionosphere's chemical (Subsection 2.3.1) and physical (Subsection 2.3.2) aeronomy on which ANIMo is partially based. Section 2.4 is dedicated to the layered structure of the ionosphere, and it is followed by the chapter summary.

2.1 From the Sun to the Earth

2.1.1 The Sun

In our solar system, the Sun makes up 99% of the total mass. Its radius is on average 696000 km, whereas the radius of our planet is 6371 km, more than 100 times smaller. In terms of volume the comparison is even more striking: 1 million Earths are required to form the volume of our Sun. The size difference is not the only impressive characteristic: the Sun radiates a tremendous quantity of energy, close to 4×10^{33} erg s^{-1}. This energy is primarily the result of nuclear fusion processes in which hydrogen is converted in helium. The fusion takes place in the zone called the core with estimated temperatures reaching approximately 1.57×10^{7} K. Surrounding the core, in an onion-like structure, there are a series of layers with different features (the radiative and convective zones). The visible surface of the Sun is called the photosphere. Here the temperature drops to 6000 K. There are regions of the photosphere which look darker from the Earth and that are relatively colder (4000 K) than the surrounding temperatures. They are called Sunspots and they are particularly useful for understanding space weather. Their number, positions and arrangements are strongly related to the solar magnetic field. Despite the fact that their formation is not entirely well understood, scientists can use them as indicators of solar activity. There are, in fact, periods when the Sun is quiet and periods in which violent phenomena are more likely to occur. These phases are called Solar Minimum and Solar Maximum respectively; together they constitute a Solar Cycle which lasts around 11 years.

2.1.2 Solar emissions and the Interplanetary Magnetic Field (IMF)

Coronal Mass Ejections (CMEs) can be four times more probable during Solar Maxima [*Tascione*, 1994]. CMEs are strong solar events where a huge amount of plasma and radiation leaves the Sun vigorously through the solar atmosphere (corona) and can reach the Earth and perturb the geomagnetic field and the ionosphere. Solar flares and eruptive prominences are other famous coronal events, in which very energetic explosions generate a short-lived burst of electro-magnetic radiation. The Sun normally releases energy, partly in the form of electromagnetic radiation and partly corpuscular flux. The solar electromagnetic radiation covers a wide range of wavelengths from X-rays down to radio waves. A generally accepted and widely used indicator of the solar activity level is the measurement of solar flux on the specific radio wavelength of 10.7 cm (which corresponds to 2800 MHz). It is called the *F10.7* index (sometimes the Convington Index) and it has a good positive correlation with X-rays, extreme ultraviolet, and ultraviolet fluxes. There is also a formula that provides a mathematical relationship between the *F10.7* index and the number of sunspots observed [*Leitinger et al.*, 2005]. However, the Solar constant (1370 W/m^2) is the conventional value of the amount solar energy that eventually reaches the surface of the Earth.

As previously mentioned, particles are also continuously emanating from the Sun. This emission is plasma from the corona that expands radially towards the 'free' space surrounding the Sun; this is the solar wind. This stream of particles is strongly embedded in an important element of the Sun-Earth environment: the IMF. The IMF is a weak magnetic field and it is generated by the Sun through a process that is called the solar dynamo. Because of solar rotation, the IMF expands radially and its field lines assume a spiral shape. The solar wind flows through the IMF and accelerates or decelerates according to the orientation of the IMF's lines. When entering and exiting from the Sun, these lines produce regions on the solar surface with different polarities called magnetic sectors. Generally the orientation of the IMF is another indicator of the level of the solar activity. The IMF also tends to be more chaotic when the Sun is more active. Between IMF lines with opposite directions, it creates the interplanetary sheet current. This current spreads all around the Sun within the interplanetary medium in a typical Parker spiral shape. Because the solar magnetic field is rotating the current sheet waves, similar to a ballerina's skirt, and it is crossed by the Earth's orbit. The position of the Earth with respect to the sheet is very important from a space weather point of view, the reason will be explained below.

2.1.3 The geomagnetic field

The Earth also has its own magnetic field which is often approximated as a magnetic dipole, but if it is observed from an appropriate distance it has a pronounced comet-like shape with the tail pointing away from the Sun. This is due to the solar wind that 'blows away' the geomagnetic field. The result of this interaction is called the magnetosphere. The plasma coming from the Sun towards the Earth is mostly slowed down and deflected by this magnetosphere but a small part actually passes through. The orientation of the IMF can considerably vary because of CMEs. The IMF arrangement can modify the architecture of the geomagnetic field and hence increase the portion of entering particles. This happens when the vertical IMF component (B_z) is negative for prolonged periods, in other words when the Earth is beneath the current sheet.

In these conditions the IMF lines are oriented in a way that allows them to be combined with the geomagnetic field lines. The interaction deforms the inner structure of the magnetosphere, changes the currents systems, creates magnetic recombination processes, and permits the solar particles to reach the atmosphere. One of the indirect effects is the possibility, normally in high-latitude regions, to observe magnificent auroral events. These heavy perturbations of the geomagnetic field can also bring less pleasing effects. 'Geomagnetic storm' is the general term that is attributed to strong magnetosphere disturbances. Depending on their intensity, geomagnetic storms can heavily affect human activity for example by damaging power grids or impeding communications.

There are various indices that describe the level of geomagnetic perturbation. The most famous is the Disturbance Storm Time index (D_{st}). D_{st} basically measures the intensity of the ring current which increases during magnetic storms. The *K* (from Kennziffer=index) indices are widely used

for measuring the geomagnetic activity. It is collected every 3 hours and its values go from 0 to 9 in a quasi-logarithmic scale. The K_p (planetarische Kennziffer) index is the global version of the K index and is calculated by combining measurements from various ground-based magnetic observatories. It is quite often used in modelling but its logarithmic nature makes the K_p inappropriate if daily values are needed. The introduction of the A_p (daily equivalent planetary index) solves this problem. It is the daily average of a 3 hour interval value (a_p) that is directly derived from the K_p index [*Davies*, 1990]. Also from data series of geomagnetic indices it is possible to appreciate a certain periodicity which corresponds to the 11-year solar cycle. There is indeed a relationship between solar and geomagnetic activity (with a little delay in the latter), demonstrated by a strong correlation between number of sunspots and K_p values [*Matsushita and Campbell*, 1967].

Another important parameter which is very useful for ionospheric modelling is the magnetic dip angle. As previously mentioned, the geomagnetic field can be compared to a magnetic dipole. In order to simulate plasma motions in the ionosphere, the awareness of the inclination of the magnetic field lines is crucial. The geomagnetic dip gives this kind of information. It is the angle formed by a compass needle (designed to move vertically) and the Earth's horizon. When the magnetic field lines are parallel to the surface, such as at the equator, the dip angle is equal to 0°. At the magnetic pole, where the field lines are vertical, the dip angle assumes values close to 90° or -90°. The angles are positive in the northern hemisphere and negative in the southern. The geomagnetic field is actually more complex than a magnetic dipole. The positions of the magnetic poles are not aligned on the same axis and the magnetic equator is wavier than a normal sphere intersection. Furthermore the configuration of the field is constantly evolving and this implies, for example, the change of position of the magnetic poles on the Earth surface. The International Association of Geomagnetism and Aeronomy (IAGA) takes care of preserving and updating a standard mathematical description of the field called International Geomagnetic Reference Field (IGRF) which is now the '12th generation'. Thanks to this tool, geophysicists and modellers can retrieve estimates of dip angles for any point on the Earth.

2.2 The ionized air

The ionosphere is "*the part of the earth's upper atmosphere where ions and electrons are present in quantities sufficient to affect the propagation of radio waves*" [*Rishbeth and Garriott*, 1969]. The latter part indicates why it is so often studied and, in a way, how it was discovered. Although the presence of a conducting layer in the upper atmosphere had already been proposed by Carl Friedrich Gauss, in 1901 Guglielmo Marconi was the first to exploit it by sending a radio signal over the Atlantic. Because of its key role in radio communication, in the last century the ionosphere was a frequent topic of scientific studies and experiments. Nowadays scientists are using this rich abundance of

historic knowledge together with modern tools in order to understand how the ionosphere behaves and how its behaviour affects modern communication and navigation systems.

Essentially, the atmosphere is a gas shell that wraps around the earth and it can be described as a stratified structure containing different vertical gradients. When taking temperature into consideration from sea level to almost 1000 km, the atmosphere is so divided: troposphere, stratosphere, mesosphere, thermosphere, and the exosphere. The exosphere slowly merges with outer space. The atmosphere can also be divided into a neutral atmosphere and an ionosphere. In this case the classification criterion is the presence of free charged particles mainly due to solar photoionization. The ionosphere starts at around 70 km above sea level and it approximately includes the layers above the stratosphere. Its extension and its structure can actually vary according to daily, seasonal and solar activity fluctuations.

The composition of neutral species is a key element in determining structure of the ionosphere. The major neutral constituents are N_2, O_2, and monoatomic oxygen, there are then minor neutral species He, Ar, H and N. The heavier neutral constituents are more concentrated in the lower part of the ionosphere. In general, their densities quickly drop as altitude increases (Figure 1).

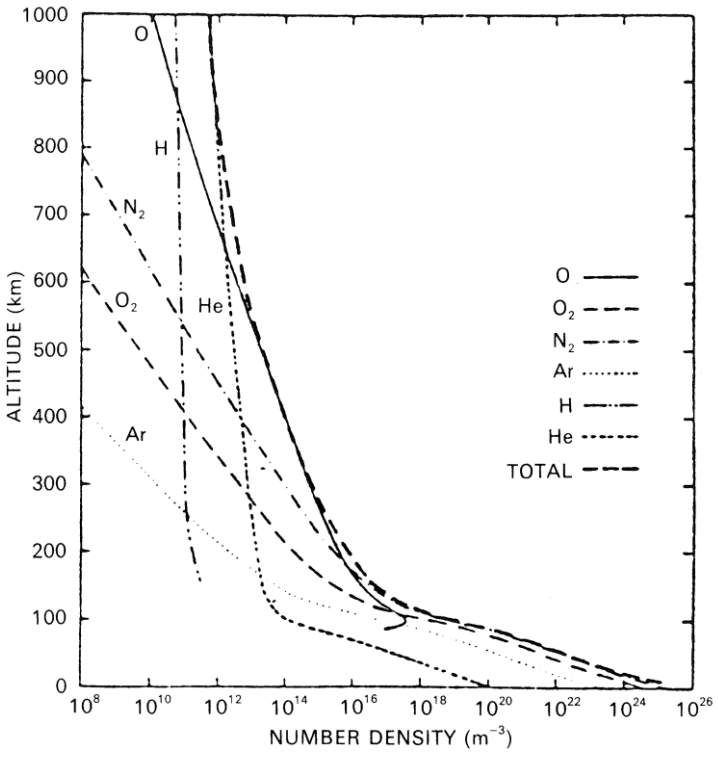

Figure 1. Vertical density profiles of the neutral species [*Hargreaves*, 1992].

At the same time, solar radiation travels through the atmosphere and gets increasingly absorbed and hence gradually loses intensity as it reaches the surface of the earth. The combination of these effects produces an enhancement of ion and electron production where the atmosphere is

populated enough by neutral species and the intensity is still strong enough to induce photoionization. Sydney Chapman in 1931 introduced and implemented this concept, which is probably the first attempt at an ionospheric model. Thanks to this theory he was able to reproduce the typical shape of the vertical profile of the electron density for an idealized atmosphere: the famous Chapman Layer.

2.3 Aeronomy

The following section applies to material adapted from various sources [*Davies*, 1990; *Hargreaves*, 1992; *Rees*, 1989; *Rishbeth and Garriott*, 1969; *Tascione*, 1994].

2.3.1 Chemical aeronomy

From a chemical point of view the principal ionospheric photoionization processes can be described as follows:

$$O + h\nu\ (< 911\ \text{Å}) \rightarrow O^+ + e^- \tag{2.1}$$

$$N_2 + h\nu\ (< 796\ \text{Å}) \rightarrow N_2^+ + e^- \tag{2.2}$$

$$O_2 + h\nu\ (< 1026\ \text{Å}) \rightarrow O_2^+ + e^- \tag{2.3}$$

Where h is Plank's constant, it is multiplied by ν that is the frequency of the electromagnetic wave of the absorbed photon. In the equation, it represents the energetic trigger (quantum of energy) of the ionization. Each species has different and specific ionization threshold energies and wavelengths. Another important process in the formation of the ionosphere is called photodissociation:

$$O_2 + h\nu\ (< 2422\ \text{Å}) \rightarrow O + O \tag{2.4}$$

This is particularly relevant for the production of monoatomic oxygen that can be subject to a photoionization. If the involved photon has enough energy, there is the possibility that a combination of ionization and disassociation takes place:

$$O_2 + h\nu\ (< 662\ \text{Å}) \rightarrow O^+ + O + e^- \tag{2.5}$$

$$N_2 + h\nu\ (< 510\ \text{Å}) \rightarrow N^+ + N + e^- \tag{2.6}$$

The wavelengths, which are reported the equations (2.5) and (2.6), are the ionization threshold of the relative ions in their ground state. There is also the possibility to have more than one ionization reaction on the same species, therefore more than one excited state. However, further ionizations require higher amounts of energy.

Photoionization is not the only type of reaction and in particular it is not the only source of energy to create charged particles. There is in fact the possibility to have ionization due to the collision between neutral species and ionised particles coming from the Sun or from auroral events:

$$N_2 + e^-_p \rightarrow N_2^+ + e^-_p + e^-_s \quad (2.7)$$

$$N_2 + e^-_p \rightarrow N^+ + N + e^-_p + e^-_s \quad (2.8)$$

$$O_2 + e^-_p \rightarrow O_2^+ + e^-_p + e^-_s \quad (2.9)$$

$$O_2 + e^-_p \rightarrow O^+ + O + e^-_p + e^-_s \quad (2.10)$$

$$O + e^-_p \rightarrow O^+ + e^-_p + e^-_s \quad (2.11)$$

Where e^-_p is the solar or auroral electron (primary); and e^-_s is released from the neutral (secondary). Primary electrons are also able to provoke dissociation:

$$O_2 + e^-_p \rightarrow O^* + O + e^-_p \quad (2.12)$$

$$N_2 + e^-_p \rightarrow N^* + N + e^-_p \quad (2.13)$$

The asterisk indicates the atom in an excited state. The production of ions is also given by charge rearrangements between ions themselves:

$$O^+ + N_2 \rightarrow NO^+ + N \quad (2.14)$$

$$O^+ + O_2 \rightarrow O_2^+ + O \quad (2.15)$$

$$O_2^+ + N_2 \rightarrow NO^+ + NO \quad (2.16)$$

$$N_2^+ + O \rightarrow NO^+ + N \quad (2.17)$$

$$N_2^+ + O_2 \rightarrow O_2^+ + N_2 \quad (2.18)$$

The above equations are called charge-transfer or atom-ion exchange reactions. It is possible to have other kinds of charge reorganizations, for example between atoms (i.e. not necessarily in the presence of an ion) or due to a simple transfer of charge between species.

Electrons and ions are unstable species and they tend to recombine in order to return to a neutral state. These are called recombination reactions and they are responsible for the disappearance of ions and electrons. For this reason they are often associated with the concept of ion or electron loss. The most relevant loss reactions are given below:

$$O^+ + e^- \rightarrow O + h\nu \tag{2.19}$$

This type is called radiative recombination because it causes the release of a photon. Another important series of reactions can be called dissociative recombination reactions:

$$O_2^+ + e^- \rightarrow O + O \tag{2.20}$$

$$N_2^+ + e^- \rightarrow N + N \tag{2.21}$$

$$NO^+ + e^- \rightarrow N + O \tag{2.22}$$

Electrons can also connect to neutral species in a reaction known as attachment:

$$O_2 + e^- \rightarrow O_2^- \tag{2.23}$$

It is important to note that the equations reported here are only a few of the total reactions that take place in the ionosphere and more species can be included.

2.3.2 Physical aeronomy

Ion and electron production and loss are often summarized in a continuity equation for charged particle densities. From a physics point of view, the equation enables the description of the behaviour of the ionosphere, respecting mass and electric charge conservation. The continuity equation for electron density is:

$$\frac{\partial N_e}{\partial t} \rightarrow Q - L(N_e) - \mathrm{div}(N_e v) \tag{2.24}$$

This equation shows that it is possible to calculate the rate of change in time t of the electron density N_e by quantifying processes of electrons: production Q and loss L and the transport term $\mathrm{div}(N_e v)$.

The production term Q depends on a series of different factors. First of all, according to the explained photoionization reactions, Q is proportional to the density of neutral species. Furthermore, it is influenced by the intensity of the solar flux. This translates into a variation of the production related to day/night and seasonal changes. The solar activity can then modify the importance of Q. It has to be also taken into account that the solar rays are absorbed while they travel through the ionosphere. The rate of absorption for a particular ray depends on its geometry (how long it travels and with which inclination) and on the concentration of the neutral species which are responsible for reducing the solar intensity. The loss term L is related to the electron density N_e. It is also influenced by temperature, which in turn depends on the solar irradiance and on collision processes between ionospheric particles. There is an additional term to the described continuity equation (2.24). The ionosphere is indeed ruled also by processes of charged particles

movement, especially in its upper layers. It is normally referred to as the transport term $\mathrm{div}(N_e v)$ and indicated as a loss; div indicates the divergence operator and v is the mean velocity.

Ions and electrons travel in the ionosphere principally by three transport processes: plasma diffusion, electromagnetic drift and neutral wind [*Davies*, 1990]. Plasma physics is fundamental to these processes. As previously described, plasma is a gas that contains ionized particles so it cannot be physically treated simply as a fluid and it is necessary to take into account its charged behaviour. One of the most important reasons for its deviation from simple fluid behaviour is the interaction with the magnetic and electric fields of the Earth. These fields modify the motion of electrons and ions which generates movement phenomena such as electromagnetic drifts.

Atmospheric tides are waves at the global-scale generated by the thermal action of the Sun and the gravitational action of the Moon. These tidal forces produce a motion of air principally in the horizontal direction. Travelling across the magnetic field $\vec{\boldsymbol{B}}$, the motion of air develops an electric current system and relative an electric field $\vec{\boldsymbol{E}}$ at an ionospheric level. The presence of the electric field is particularly important because it forces charged particles to cross the geomagnetic field lines and so explains various ionospheric plasma upward motions. In general this process is called electromagnetic drift or ExB drift; because electric and magnetic fields are approximately perpendicular to each other in the ionosphere, it is possible to write the ExB velocity as follows:

$$v_{drift} = \frac{\vec{\boldsymbol{E}} \,\mathrm{X}\, \vec{\boldsymbol{B}}}{\left\|\vec{\boldsymbol{B}}\right\|^2} \tag{2.25}$$

The resultant velocity will be normal respect to $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$.

In the Earth's ionosphere, plasma is also affected by vertical diffusion. The formula that describes this transport process is basically a balance between the downward motion due to the Earth's gravity and the upward motion due to a vertical gradient of pressure. The vertical diffusion velocity can be calculated by the following equation:

$$v_{diffusion} = -D_a(z)\, sin^2 I \left(\frac{1}{N_e} + \frac{\partial N_e}{\partial z} + \frac{mg}{2k_B T}\right) \tag{2.26}$$

D_a is the ambipolar diffusion coefficient, z is the altitude, I is the dip angle, N_e is the electron density, m is the particle mass, g is the gravity acceleration, k_B is the Boltzmann constant and T stands for temperature. The coefficient D_a depends on the altitude, the formula to calculate D_a is:

$$D_a = \frac{k_B(T_i + T_e)}{m_i \nu_{in}} \tag{2.27}$$

T_i and T_e are ion temperature and electron temperature respectively, m_i the mass of the particle mass and ν_{in} is the ion-neutral collision frequency which depends on the density of the studied neutral species.

Under the diffusive motion, the particles are not able to cross the magnetic field lines. From (2.26), it can be seen that on the equatorial zone the diffusion coefficient becomes void or very small, and that it is more important at higher latitudes and reaches the maximum at the magnetic poles [*Davies*, 1990]. The last important transport process is the neutral wind, or thermospheric wind. It is caused by a pressure difference due to a temperature day/night excursion that arises in the upper atmosphere. The effect is a horizontal movement of neutral species from the day side to the night side. Plasma is pushed by this lateral motion but not across the magnetic field. Charged particles slide along the fields' lines and they are lifted upwards when the horizontal wind is blowing toward the equator and they are dragged down when this happens in the Polar Regions. The vertical component of the plasma velocity due to horizontal wind can be calculated as follows:

$$v_{neutral\ wind} = V_{wind}\ sinI\ cosI \tag{2.28}$$

In this case, V_{wind} is the speed of the neutral wind and I is always the magnetic dip angle. The formula indicates that the transport effect of the neutral wind on plasma is greatest at middle latitudes [*Davies*, 1990].

All the concepts explained so far provide the basics for understanding the ionosphere and its behaviour; however they are only part of a plethora of involved phenomena. The complexity of the ionosphere is also due to the fact that certain processes are more likely to happen in certain periods and in certain locations. There are for example effects that take place in the equatorial region (the Appleton effect) and others that manifest only at the polar latitudes (plasma patch convection). This makes studying the processes and their interactions more difficult.

2.4 The ionospheric layers

Vertically the ionosphere presents a generalised structure however, when looking at an electron density vertical profile it is possible to identify a series of layers. Each layer presents its own characteristics and has been given an internationally recognised letter. Starting from a height of 50-60 km to 90 km above sea level there is the D region. The D region is characterized by a weak electron density (10^8-10^{10} m^{-3}) and it is absent at night. N_2, O_2, and NO are the most abundant neutral particles. Photoionization of NO by solar X-rays is the most important reaction of ion production for this layer. A special feature of this layer is the presence of negative ions produced by electron attachment reactions (as described in Reaction (2.23)). They are then destroyed by photo-detachment, associative detachment and mutual neutralization reactions. Above the D

region, between 90 and 120 km, there is the E region. Here the N_2, O_2, and NO are still the major neutral species, O_2^+, and NO^+ are the most common ionized particles (several 10^{11} m^{-3}). O_2^+ is easily generated by the photoionization (2.3) by X-ray and UV, NO^+ is instead the product of a fast charge rearrangement (2.14), (2.16), (2.17). Dissociative recombination (2.20) (2.22) is the major factor of loss of charged particles. The transport element is not really effective; in fact the E layer dynamics can be described with a simplified continuity equation where the loss term is proportional to the squared ion/electron density:

$$\frac{\partial N_e}{\partial t} \rightarrow Q - k_L(N_e)^2 \tag{2.29}$$

Where k_L indicates a generic constant. This condition is called photochemical equilibrium and can be fairly accurately reproduced using the Chapman model. The E layer diminishes but persists during the night. In this region it is also possible to notice a sporadic E-layer, also known as E_s. E_s is an irregular layer formed by localized clouds of plasma; they mostly appear during the day time with little seasonal variation. The E_s-layer develops around 100-120 km as a thin ionized stratus of height extent about 1 km [*Barclay*, 2003]. It influences radio communication greatly because sporadic E enables long distance signal propagations that are otherwise not possible to occur.

Above the E region, the F layer goes from 170 km up to over 600 km. It can be divided approximately at 200 km in two different layers, F1 and F2. F1, which tends to disappear during night, can be considered as a transition region between the E-layer and the upper part of the ionosphere. The major neutral species are N_2, O_2, and O, the ion particles are O_2^+, NO^+, and O^+ with densities of several 10^{11} - 10^{12} m^{-3}. The photoionization of O_2 and N_2 (reactions (2.4) (2.6)) are the most important sources of ions. One of the differences between this region and the E layer is its high level of O^+. Its direct recombination (radiative recombination (2.19)) is very slow hence it mostly does not take place. O^+ is lost by a chain of reactions that start with atom-ion interchange with O_2 and N_2 ((2.14)(2.15)) followed by dissociative recombination (reactions (2.20)(2.22)). Although the F1-layer is characterized by these O^+ dynamics, it is still ruled by the photochemical equilibrium. In order to have a non-Chapman type layer, it is necessary to move up to 200 km. Here the F2-layer begins. It extends to 600 km of altitude, with a peak of electron density at 250-300 km. This region represents the highest concentration of charged particles; it persists overnight, and therefore has, a very important role in space communication. The ion density reaches several 10^{12} m^{-3} and essentially consists of O^+. The ionization mechanisms are the same as those for the F1-layer except that they are magnified in the F2-layer. In fact, although the neutral species N_2, O_2, and O, are still present, their densities rapidly diminish along the profile. This affects the rate of recombination which can be now assumed only proportional to the charged particle density. The deviation from the photochemical equilibrium is also due to the activity of ionospheric

transport processes which become more important due to particular physics conditions. Above the F2 peak the electron density decreases with altitude. At 400 km there is a significant concentration of H^+ and He^+ which require the explanation of further chemical processes that will not be discussed in this report. The presence of these light ions and the dominance of transport processes are the most significant features of the topside region of the ionosphere (600-1000 km).When the concentration of H^+ and He^+ becomes greater than the atomic oxygen ion one, a fully ionised region called the plasmasphere or protonsphere begins.

Figure 2, taken from the work of [*Hargreaves*, 1992], shows typical vertical profiles of electron density at a generic mid-latitude location. It summarizes very well both the aforementioned ionospheric structures and their variability.

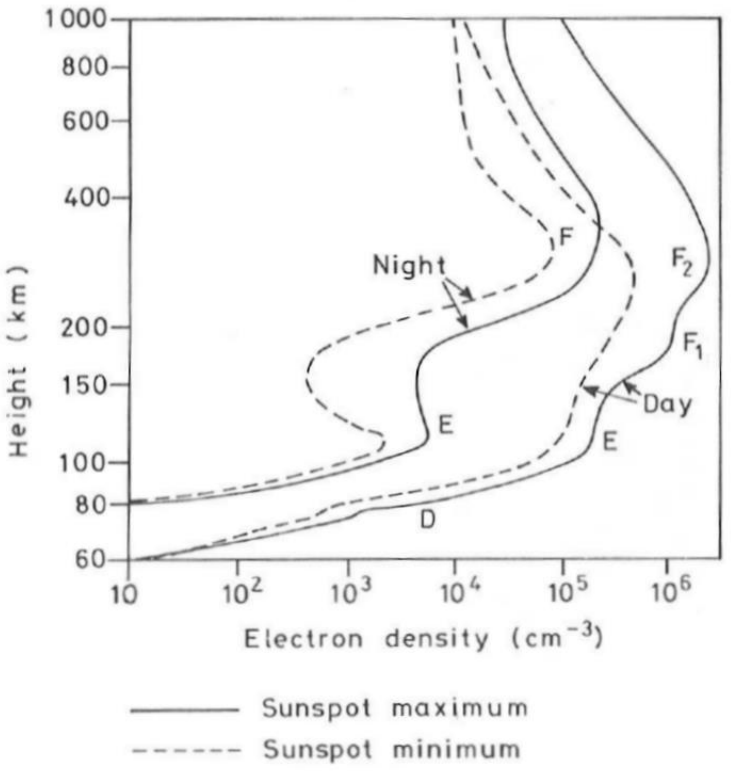

Figure 2. Electron density vertical profiles from a mid-latitude location in different diurnal and solar activity circumstances [*Hargreaves*, 1992].

The graph enables us to appreciate diurnal changes and changes related to the solar activity. The solid lines are the electron density profiles related to high solar activity and the dashed lines to low solar activity. Note that the changes are not simply related to the density; the layer's shape is also very much affected.

Summary

The Sun-Earth system and the terrestrial upper atmosphere were briefly described in this chapter. The latter was structured in order to explain, at the same time, concepts that have been useful during the project: the definition of the principal processes, the meaning of different parameters and in general the characterization of the ionospheric medium. For this reason, some of the topics will be propose again in the following chapters. In particular, Chapter 5 reports how the principal ionospheric dynamics have been implemented in the model ANIMo. For example, photo-ionization and ambipolar diffusion transportation will be further expanded.

The next chapter continues the background review by introducing some of the techniques used for measuring the ionosphere and its features. It focusses on the sources of observations that were adopted during the project.

Chapter 3 Observational techniques and data sources

Abstract

GPS was designed to provide location and timing. GPS satellites signals are affected by the presence of the ionosphere; in particular they are subject to delays that are proportional to the amount of electrons along the ray-paths. The electron density can be estimated from a dual-frequency observation of signal delays by extracting the dispersive elements and applying a calibration.

The ionosonde is an instrument that was specifically built to measure features of the ionosphere. It sends a series of pulses at different frequencies and records what is reflected by the upper atmospheric structures. From its observations it is possible to measure the altitude of the layer peaks and the electron density at such altitudes.

Another well-known instrument is the Incoherent Scatter Radar (ISR). It exploits back-scattered signal coming from the electrons and ions motion in the ionosphere. It can provide a wider range of measurements with respect to the ionosonde, however it is more expensive to build and operate.

Introduction

A vast amount of data were used during the project period to underpin the development of ANIMo and the data ingestion scheme ANIDAS. GPS measurements have been fundamental as they represent the observation contribution in ANIDAS (Chapter 7). Ionosonde and ISR measurements have been constantly adopted to assess the accuracy of the presented approaches especially with regards to the absolute peak density and the vertical resolution. ANIMo's and ANIDAS's final validation tests (Chapter 6 and Chapter 8) are based on comparisons against these instruments. In addition, ISR temperature measurements were fed into ANIMo (Section 6.3) to test its sensitivity and understand whether this input makes a difference in the simulation results. Drift vertical velocities from the same instrument were investigated in order to ameliorate and correct ANIMo physics of the plasma transportation processes. Other data, such as A_p and *F10.7* indexes that have been introduced in the previous chapter, are fundamental to simulate different geomagnetic and solar activity situations in ANIMo.

In this Chapter, Section 3.1 describes the GPS system (Subsection 3.1.1) and how it can be used to provide ionospheric measurements (Subsection 3.1.2). Section 3.2 introduces the theory of

ionospheric sounding by describing the operation of the ionosonde (Subsection 3.2.1). Similarly, Section 3.3 is dedicated to the ISR and the physical concepts (Subsection 3.3.1) behind its measurements (Subsection 3.3.2). Section 3.4 reports the data sources used in this project; their references are added in the Acknowledgements.

3.1 GPS for ionospheric measurements

3.1.1 GPS and positioning

The first navigation and positioning satellite system was the NAVigation Satellite Time and Ranging (NAVSTAR) Global Positioning System (also called GPS). It was developed in United States (US) in the 1970s by the Department of Defence (DoD) and it has been operative since the 1980s. It is a constellation of 24 to 32 satellites which orbit at an altitude of approximately 20000 km. They are divided into six fixed orbital planes which have an inclination of 55° with respect to the equatorial plane. A satellite takes a little less than 12 hours to complete its orbit, and from a single fixed location on the Earth it is visible for approximately five hours provided no obstacles are present. These conditions enable a GPS user to always rely on an optimal satellite arrangement. The GPS satellites broadcast two frequencies – L1=1575.42 MHz that is modulated by a public Coarse/Acquisition code (C/A) and an encrypted precision code (P or Y), and L2=1227.60 MHz modulated only by (P).

The navigation message broadcasted by each satellite carries information about the position of the satellite and the time when the signal was transmitted. The GPS receiver collects this message from multiple satellites and, by assuming that the message travelled at the speed of light, it estimates the distance (pseudo-range) travelled by each signal. The position of the receiver is calculated through a trilateration. At least four satellites are necessary in order to perform this calculation because of the poor accuracy of the receiver clock.

The satellite signal is affected by various propagation errors that may cause significant errors in the positioning calculation. The biases can be analysed by expanding the GPS pseudo range and carrier phase measurements [*Jakowski*, 1996]:

$$P = \rho + c(dt - dT) + d_I + d_T + d_M + h_s + h_r + \epsilon \quad (3.1)$$

$$\Phi = \rho + c(dt - dT) - d_I + d_T + d_M + \lambda_s N + h'_s + h'_r + \epsilon' \quad (3.2)$$

where, P is the pseudo-range, Φ is the measured carrier phase, ρ is the real distance (range), c is the speed of light, dt and dT are the clock offsets of the receiver and the satellite respectively, d_I is the error due to the presence of the ionosphere, d_T is instead relate to the troposphere, d_M is the error due to multipath effects, h_s and h_r are the satellite and receiver hardware dispersive

components, λ_s is the wavelength of the signal, N is the carrier phase ambiguity, ϵ is the residual error.

3.1.2 Measuring ionospheric electron density with GPS

GPS data can be used for monitoring the ionosphere and its behaviour. The measurement of the integrated electron density, which is defined as the Total Electron Content (TEC), can be retrieved by analysing the ionospheric delay. TEC is an integrated measurement of the amount of free electrons contained in a hypothetical column of 1 m^2 cross-section built along an ionospheric path and it is expressed in TEC Units (1 TECU = 10^{16} m^{-2}). In this case, the path refers to a segment of ionosphere between a given satellite (transmitter) and receiver, and it can be defined by the following formula:

$$TEC = \int_{TX}^{RX} N_e \, ds \tag{3.3}$$

TX and RX are the positions of the transmitter and of the receiver respectively, N_e is the electron density and s is the ray path. The measurement can be estimated by exploiting the influence of the electron density on the propagation of the GPS signal. The velocity of the latter, being electromagnetic wave propagation, is affected by the refractive index of the medium through which it is traveling. The refractive index of a certain medium can be defined as the ratio of the speed of light in vacuum to the velocity of the electromagnetic wave in that medium. The refractive index n^2 of the ionosphere is described by the Appleton-Hartree equation:

$$n^2 = 1 - \frac{X}{1 - iZ - \frac{Y_T^2}{2(1 - X - iZ)} \pm \sqrt{\frac{Y_T^4}{4(1 - X - iZ)^2} + Y_z^2}} \tag{3.4}$$

Where the term Z considers the electron-neutral collisions, Y is instead related to the presence of the geomagnetic field and X is equal to:

$$X = \frac{N_e e^2}{\varepsilon_0 m_e \omega^2} \tag{3.5}$$

Where e is the charge of an electron, ε_0 is the permittivity of the free space, m_e is the mass of the electron, ω is the angular wave frequency. By assuming the absence of geomagnetic field ($Y = 0$) and of electron-neutral collisions ($Z = 0$), equation (3.4) can be reduced to:

$$n^2 = 1 - X \tag{3.6}$$

That can be rewritten in these terms:

$$n^2 = 1 - k\frac{N_e}{f^2} \tag{3.7}$$

Where f is the signal frequency and k is:

$$k = \frac{e^2}{4\pi^2 \varepsilon_0 m_e} = 80.6 \text{ m}^3\text{Hz}^2 \tag{3.8}$$

In order to retrieve the electron density along the signal path, it is possible now to perform differences between the pseudo-range equations (3.1) of the two GPS frequencies (L1 and L2). By considering only the dispersive elements of equation (3.1) and assuming that the multipath delay is not frequency dependent, the differences will be [*Mannucci et al.*, 1999]:

$$P_{L2} - P_{L1} = I_p(1/f_{L2}^2 - 1/f_{L1}^2) + b_r + b_s \tag{3.9}$$

Where b_r and b_s are the biases from the hardware dispersive components of receiver and satellite respectively, I_p is an ionospheric related parameter directly proportional to the unknown TEC value:

$$I_p = 40.3 \int_{TX}^{RX} N_e \; ds = 40.3 \; TEC \tag{3.10}$$

The differences approach can be similarly implemented for the carrier phase measurement:

$$\Phi_{L1} - \Phi_{L2} = I_p(1/f_{L2}^2 - 1/f_{L1}^2) + (\lambda_{L1} N_{L1} - \lambda_{L2} N_{L2}) + b'_r + b'_s \tag{3.11}$$

Although equations (3.9) and (3.11) are actually measuring the same delay, they have a significant difference. The one related to the pseudo-ranges (3.9) provides an absolute TEC values. Its estimation is however very noisy, especially when the ray path is traveling at low elevation angles and is subject to multipath. On the other hand, because of the presence of the carrier phase ambiguities N_{L1} and N_{L2}, equation (3.11) is not able to provide TEC values in absolute terms. The carrier-phase measurement is nevertheless cleaner with respect to the pseudo-range one, even at low elevation angles. The combination of the two, by means of a least square error approach, can provide the calibration of the TEC, with differential code biases still corrupting the TEC observation. Figure 3 by *Jakowski* [1996] shows the aforementioned.

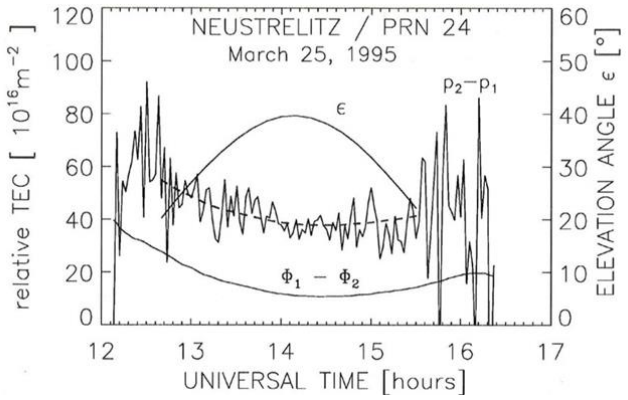

Figure 3. Differential pseudo-ranges and carrier phases relative to satellite PRN24 reproduced over Neusterlitz (Germany) on the 25th of March 1995. The dashed line shows how the least square method over imposed the two differentials in order to have a calibrated measurement [*Jakowski*, 1996].

If the remaining biases can be corrected for, the retrieved TEC value is also known as calibrated Slant TEC (STEC) as it is related to the inclination of the satellite ray path. By projecting the STEC to a vertical profile, this measurement can be converted to Vertical TEC (VTEC). The operation involves an approximation that loses its validity as the ray path tends towards the horizon. More accurate methods enable accumulation and interpretation of multiple STEC measurements and convert them to vertical electron density profiles and therefore produce TEC maps. These approaches lead to sophisticated methods called ionospheric tomography.

3.2 Ionospheric sounding

The ionosonde is a very well established instrument. Because of its relative low cost of installation and operations, a large number of these devices are spread around the world to continuously monitor the ionosphere.

Several textbooks [*Davies*, 1990; *Hargreaves*, 1992; *Rishbeth and Garriott*, 1969] were consulted for preparing this section.

3.2.1 Ionosonde principles

The fundamental principle of the sounding technique, introduced by *Breit and Tuve* [1925], is to send a pulsed radio signal, with known frequency, vertically and measure the time that passes before its echo it is received back. Considering the simplified Appleton-Hartree equation (3.6), where geomagnetic field and collisions are neglected, and the definition of X from Equation (3.5) it is possible to write:

$$n^2 = 1 - \frac{N_e e^2}{\varepsilon_0 m_e \omega^2} \tag{3.12}$$

At a fixed transmission frequency ω, as the electron density increases the refractive index n^2 becomes smaller. When $N_e e^2/\varepsilon_0 m_e = \omega^2$ the energy transported by the pulse is reflected back

toward the Earth surface. The concept of the ionosonde is to send a number of different pulses by sweeping along increasing frequencies and then register the delay of their echoes. The delay (or time of flight) can be converted to the altitude where the reflection took place. The latter is called virtual height h' from which it is possible to estimate the real height of reflection h_m. This is done by taking into account that the transmitted radio pulse does not travel at the speed of light in a vacuum. By reporting each frequency against the virtual height, the ionosonde produces the ionogram (Figure 4).

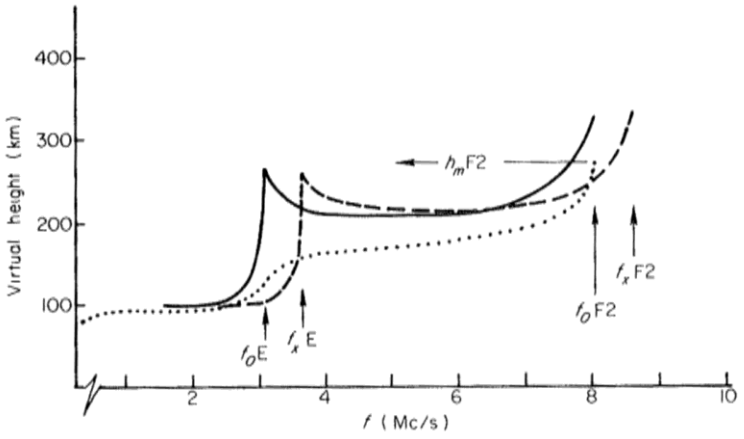

Figure 4. Idealized ionogram. The virtual height is plotted against the frequency (Mc/s is MHz), the solid and dashed lines refer to ordinary and extraordinary reflected wave respectively. The dotted line is the real height versus the frequency [*Rishbeth and Garriott*, 1969]

Figure 4 shows an ideal ionogram taken from the work of *Rishbeth and Garriott* [1969] where two layers (E and F2) are registered by the ideal ionosonde. The solid and dashed lines refer to two different kinds of propagation: the "ordinary" and the "extraordinary" respectively. This is due to the presence of the magnetic field ($Y \neq 0$ in the Appleton-Hartree equation (3.4)) which makes the ionosphere behave as a double refracting medium. To be more precise, a transmitted radio wave splits into the ordinary and extraordinary components when entering the ionosphere; they are called characteristic waves and travel independently. Because of its polarisation, the electric field associated with the extraordinary wave accelerates electrons which travel by gyrating around the Earth's magnetic field lines [*McNamara*, 1991]. This influences the electrons' motion and therefore modifies the refractive index. The result is that the reflection of the extraordinary wave happens at higher frequency with respect to the ordinary wave by a factor that is half the electron gyro-frequency.

Figure 4 also shows that it is possible to observe the presence of spikes (where the virtual height is suddenly bigger) which are labelled by specific abbreviations (such as f_oF2). These refers to the critical frequencies f_o (or f_x for extraordinary) of different ionospheric layers (E and F2) and correspond to the maximum frequency that can be reflected by a layer; beyond the critical frequency of the F2 layer, the transmitted pulse is lost into space. Critical frequencies

(conventionally the f_o is taken as reference parameter) can be easily converted into electron density values by the following formula:

$$f_o = (80.6\, N_m)^{\frac{1}{2}} \tag{3.13}$$

Where f_o is expressed in kHz and N_m is the layer peak electron density value (in cm^{-3}). Figure 4 shows a dotted line which the plasma frequencies are plotted versus the real heights. Real heights and peak electron density of the F2 layer (h_mF2 and N_mF2 respectively) have been crucially important for this project.

3.3 Incoherent Scatter Radars (ISR)

The ISR is also a ground based instrument for ionospheric measurements. In contrast to the ionosonde (Section 3.2), the ISR can scan over the h_mF2 limit and hence it allows the study of the topside of the ionosphere. In addition, it is capable of measuring a series of other important features of the ionosphere such as temperatures, composition and drift velocities. Because it is a very expensive and rather big instrument, only about ten ISRs are operative around the world.

The works of *Beynon and Williams* [1978], *Davies* [1990] and *Hargreaves* [1992] were consulted to write the following section. Further explanations of the ISR physics principles can be found in those publications.

3.3.1 ISR principles

The ISR is based on a physics principle called Thomson scattering for which electrons are capable of scattering electro-magnetic waves (X-rays). The radar cross section associated with an electron (σ_e) is called Thomson cross section and can be estimated, for a direct backscatter, $\sigma_e = 4\pi r_e^2 = 10^{-28}\ m^2$ where r_e is the effective radius of one electron. To detect such small scattering cross section, the ISR needs to be very sensitive. About 50 years after Thomson demonstration, *Gordon* [1958] realised that there was the technology to build such a radar and this was achieved in practice the same year by *Bowles* [1958]. Gordon also predicted that due to thermal velocities in the medium, the spectrum of the scattered signal would present Doppler shifts with half power width of 0.71 Δf_e. Δf_e is intended to be the Doppler shift ($2v_T/\lambda_R$) generated by an electron moving toward the radar at the mean thermal speed v_T (where $v_T = (2k_BT_e/m_e)^{\frac{1}{2}}$) and can be, therefore, determined as follows:

$$\Delta f_e = \frac{(8k_BT_e/m_e)^{\frac{1}{2}}}{\lambda_R} = 11T_e^{\frac{1}{2}}/\lambda_R\ kHz \tag{3.14}$$

Where λ_R is the wavelength of the transmitted signal from the radar, k_B is the Boltzmann constant, T_e is the electron temperature and m_e is the mass of the electron. Considering $T_e = 1600\ K$,

$\lambda_R = 75\ cm$ ($f = 400\ MHz$), the Δf_e would be around 600 kHz. Although in theory a broad spectrum is expected (hence a very large antenna is needed), in practice the spectrum is around 200 times narrower. This makes the signal easier to detect. In addition to this, the spectrum appears to be more complex and to provide more information about the ionosphere. The unexpected features of the observed spectrum are due to the coupling between ions and electrons, for which the movements of the electrons depends on the surrounding ions. The width of the observed spectrum is so reduced by a factor of $(m_i/m_e)^{\frac{1}{2}}$ with respect to the theoretical one predicted by Gordon, where m_i and m_e correspond the ion and electron masses respectively.

The limit between the possibility of measuring the scattering of free electrons or the scattering of electrons whose motion is controlled by ions depends on the Debye length, which is the characteristic distance of plasma. Beyond the Debye length, charged particles are not affected by charges within the Debye length. In a plasma medium a charge tends to attract opposite charges and repel equal ones. The attracted charges create a screen around the initial charge that cancel its electric field effects on the surrounding particles. The Debye length can be seen as the radius of the sphere of influence of a particle; it is very important in plasma kinetics. It represents the boundary between considering particle collisions on the small scale, and plasma collective effect on a larger scale [*Hargreaves*, 1992]. The Debye length is normally defined as λ_D:

$$\lambda_D = \left(\frac{\varepsilon_0 k T_e}{e^2 N_e}\right)^{\frac{1}{2}} = 69(T_e/N_e)^{\frac{1}{2}}\, m \tag{3.15}$$

Where N_e is the electron density and e is the charge on an electron. Because T_e and N_e change with altitude, the λ_D also varies between a few millimetres to around a few centimetres. If the exploring wavelength is smaller than λ_D the radar is able to detect scattering from free electrons and the resulting spectrum corresponds to Gordon's predictions. At normal operational conditions, the transmitted wavelength is bigger than λ_D and therefore the interaction between ions and electrons needs to be considered.

Ion-acoustic and electron-acoustic waves are waves generated in the plasma by random thermal motion of the electrons. These waves, which are governed by pressure and electrostatic forces, produce the scattered signal observed by the radar. They propagate in all the directions within a wide and continuous spectrum of wavelengths. Those with a wavelength $\Lambda = \frac{1}{2}\ \lambda_R$ (where λ_R is the radar wavelength) that move along the direction of the transmitted signal at a velocity v_T will cause a quasi-coherent backscatter signal that is then measured at the ground. The Doppler shift of the scattered signal produced by an upward moving wave can be found as follows:

$$\Delta f = \text{-}2\, v_T/\lambda_R = \Lambda/\lambda_R = \text{-}F(\Lambda) \tag{3.16}$$

Where $F(\Lambda)$ corresponds to the frequency of the wave. Similarly, it is possible to calculate the Doppler shift of the down-going wave ($+F(\Lambda)$). Because there are ion-acoustic and electron-acoustic waves, the received spectrum will have four components. The frequency shift for the ion-acoustic wave is given by:

$$F^{+}(\Lambda) = \frac{1}{\Lambda}\left[\frac{k_B T_i}{m_i}\left(1 + \frac{T_e}{T_i}\right)\right]^{\frac{1}{2}} \tag{3.17}$$

In theory, this corresponds to a spectrum with two lines separated by a distance proportional to $(T_i/m_i)^{\frac{1}{2}}$. In practice, the two ion lines appear broadened (ion spectrum visible in Figure 5). This is due to a process called Landau Damping which consists in an exchange of energy between a wave and a particle when they travel in the same direction at a similar speed. If the particles are slightly slower than the wave, they accelerate receiving energy from the wave that becomes attenuated. On the contrary, if particles are slightly faster than the wave, they transmit energy to the wave which will result enhanced. Because the speed of the ion-acoustic wave is within the Maxwell distribution of thermal ion at temperature T_i, there are always more ions travelling to a slightly lower speed rather than higher speed compared to the wave. This is why the two lines are broadened and merged as shown in the ion spectrum of Figure 5.

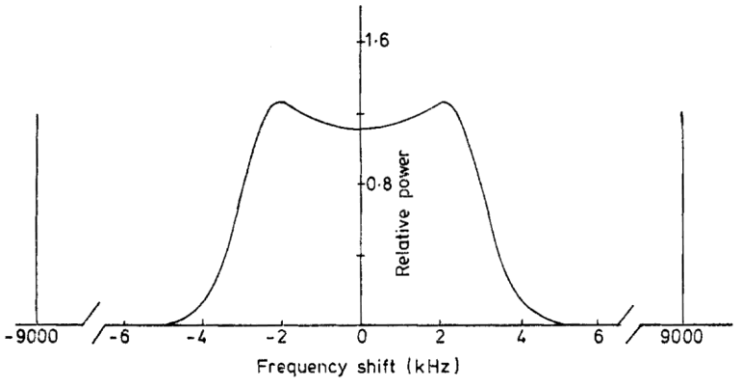

Figure 5. ISR typical spectrum [*Beynon and Williams*, 1978] (© IOP Publishing. Reproduced with permission. All rights reserved).

The frequency shift for the electron-acoustic wave is given by:

$$F^{-}(\Lambda) = f_p\left(1 + \frac{12\pi^2 {\lambda_D}^2}{\Lambda^2}\right)^{\frac{1}{2}} \tag{3.18}$$

Where f_p is the plasma frequency. Because the electron acoustic-waves travel much faster than the thermal velocities of the electron, the attenuation does not occur. The result is two sharp plasma lines whose offset is approximatively the plasma frequency of the medium (Figure 5).

3.3.2 ISR measurements

The echoes received by the ISR are quite weak, hence the data needs to be integrated. From the time delay of the received signal, it is possible to estimate the range related to the measurement and therefore a measurement profile can be easily obtained.

There are various types of measurements that can be extrapolated from the ISR spectrum (Figure 5). The most relevant for this project are the following:

- Electron density can be determined in three different ways. The first one is the most used and estimates the electron concentration of the scattering region from the total returned power. It is not an absolute method hence a calibration with another instrument (ionosonde) is necessary. The second way considers the frequency offsets of the plasma lines, which are not always strong enough to be detected. The third is based on the observation of the Faraday Effect. Although the latter is an absolute measurement, it is often calibrated against ionosonde data.
- T_e/T_i ratio can be estimated from the shape of the peak and dip of the ion spectrum.
- T_i/m_i ratio is associated with the separation of the peaks. This means that T_e and T_i can be easily calculated by modelling m_i.
- Plasma velocity is measured from the overall shift of the spectrum.

It is possible to extrapolate other interesting ionospheric parameters from the ISR spectrum that have not been mentioned here but that can be found and extensively explained, in the work by *Beynon and Williams* [1978].

3.4 Data sources

GPS measurements used during this project are RINEX datasets provided by different repositories: the Crustal dynamics data information system (NASA) (ftp://cddis.gsfc.nasa.gov/gnss/data/campaign/mgex/), International GNSS Service (IGS) (ftp://igscb.jpl.nasa.gov/pub), the Ecole Nationale des Sciences Géographiques IGS service, the University NAVSTAR Consortium (ftp://data-out.unavco.org/pub), and the GARNER GPS archive (http://garner.ucsd.edu/pub).

The ionosonde observations used in this project can be found on Space Physics Interactive Data Resource (SPIDR). This can be found at the following link: http://spidr.ngdc.noaa.gov/spidr/. Another very useful database is the Global Ionospheric Radio Observatory (GIRO) which is specifically focussed on ionosonde data (http://giro.uml.edu/).

The only ISR data source used for this project is the Madrigal database. This on-line archive consists in a wide variety of data from various instruments. The main link is: http://www.openmadrigal.org/.

The *F10.7* and A_p index, as aforementioned in Section 2.1.2 and Section 2.1.3, are daily measurements of the solar flux and geomagnetic activity respectively. For this project they are taken from the on-line database of the National Geophysical Data Center (NGDC) of the National Oceanic and Atmospheric Administration (NOAA) which is accessible through the following ftp link (ftp://ftp.ngdc.noaa.gov). Many other institutions provide this data in various form of visualization. During this project the solar and geomagnetic data section of the British Geological Survey was frequently consulted (http://geomag.bgs.ac.uk/data_service/space_weather/solar.html).

Summary

This Chapter outlines the various types of data that have been used in the present project. In particular, it focussed on the physical principles behind the measurements and the instruments that record them.

The next chapter consists in a review of the most relevant ionospheric models. Furthermore small descriptions of existing DA schemes are also included.

Chapter 4 State of the art

Abstract

The history of computer modelling of the upper-atmospheric is quite recent. The first ionospheric models were released at the beginning of the 1970s. The increasing availability of ionospheric measurements and the fast development of computer technology helped the evolution of those models over the years. In the meantime a variety of new ones have been released and progressively updated and expanded. Although some of those were already able to ingest ionospheric parameters, the first examples of a full ionospheric Data Assimilation scheme were developed only at the beginning of 2000. Several ionospheric models have been developed over the past few decades. This chapter briefly presents them by aiming to follow a chronologic order.

Introduction

The overall aim of this project is to develop an ionospheric model that is suited to DA. To achieve this, it was important to review existing models and to understand the ionospheric physics that the new model, ANIMo, needed and how to implement them to fulfil the aims of the project. Furthermore, it was necessary to select an existing ionospheric specification to use as a benchmark.

The existing ionospheric models can be broadly divided in two groups: empirical and theoretical models. The empirical ones are derived from ionospheric observations and experiments. The data used as basis to build these models can be retrieved from ionosonde, ISR, top sounders, rocket missions and other satellite instruments. For ionospheric application, empirical methods are often used to model specific features such as neutral wind or neutral composition. Theoretical methods are normally called physics based or first-principles models. They rely on mathematical equations that describe the actual electron and ion dynamics rather than statistical descriptions such as the empirical models. There are also models that lie in between the two classifications and they are called semi-empirical models. In this case the model mixes both methods.

There are also other criteria to classify ionospheric models. Two of them are the respective geographic region and the altitude range where the model is capable to provide its outputs. Another important factor is the type of output. As mentioned in the previous paragraph, some empirical models focus on specific characteristics of the upper atmosphere. Coupled models have, on the contrary, a wider capability of description which can include magnetosphere and thermosphere and their interactions with the ionosphere. These can be considered a combination of different models and they are usually physics-based.

A further useful criterion to classify the majority of ionospheric models is the list of the chemical species provided. This is quite useful as it gives an indication of the extension of the model in altitude and in terms of physics when the model is theoretical. It is also important to understand what can be the final application of the model.

Some of the above criteria are used in the Guide to Reference and Standard Ionosphere Models by the American Institute of Aeronautics [*AIAA*, 1999] and they have been adopted in this chapter to classify the most important ionospheric models.

Section 4.1 provides a review of ionospheric models, and their evolution, in order of release dates. The first subsection is dedicated to a brief explanation of some empirical models (Subsection 4.1.1), the second (Subsection 4.1.2) displays a table that collects the major ionospheric models and classify them according to the mentioned criteria. Section 4.2 examines existing DA schemes: the Ionospheric Data Assimilation Three Dimensional (IDA3D) [*Bust et al.*, 2004] in Subsection 4.2.1, the Global Assimilation of Ionospheric Measurements (GAIM) [*Schunk et al.*, 2004] in Subsection 4.2.2, the Global Assimilative Ionospheric Model (GAIM) [*Mandrake et al.*, 2004] in Subsection 4.2.3, the Electron Density Assimilative Model (EDAM) [*Angling and Cannon*, 2004] in Subsection 4.2.4 and the Multi-Instrument Data Analysis System (MIDAS) [*Mitchell and Spencer*, 2003] in Subsection 4.2.5.

4.1 Ionospheric models

The appearance of the first ionospheric model can be traced back to the beginning of the 1970's. *Nisbet* [1971], from Pennsylvania State University, proposed a semi-empirical tool for estimating electron and ion densities for the E and F regions. This was the first version of the more recent Penn state Mk III model [*Nisbet and Divany*, 1992; *Torr et al.*, 1979].

In 1973, two physics based models were created. First, the Time Dependent Ionospheric Model (TDIM) was developed by *Schunk and Walker* [1973] from Utah State University. The evolution from this model is called the USU Time-dependent model of the global ionosphere [*Schunk*, 1988; *Sojka*, 1989]. The TDIM model was further modified during the 1990's in the Ionospheric Forecast Model (IFM) [*Schunk et al.*, 1997]. IFM was then coupled to the Thermosphere Forecast Model (TFM), which is based on a previous thermosphere model developed by *Fuller-Rowell and Rees* [1980], to form the Coupled Ionosphere Thermosphere Forecast Model (CITFM) [*Sojka et al.*, 1995].

Around 1985 the Semi-Empirical Low-Latitude Ionospheric Model (SLIM) was proposed as a theoretical model for the low latitude ionosphere [*Anderson et al.*, 1987]. The second one was developed by the Air Force Research Laboratory (AFRL) and is called the Global Theoretical

Ionospheric Model (GTIM). Initially, it was dedicated exclusively to the low latitude [*Anderson*, 1973] but was later extended to a global scale [*Decker et al.*, 1994].

In 1975 the development of a further physics-based model was started in Sheffield University which became globally applicable in 1978; its name is the Sheffield University Plasmasphere-Ionosphere Model (SUPIM) [*Bailey and Sellek*, 1990; *Bailey et al.*, 1993; *Bailey et al.*, 1997].

In the late 1970's the first International Reference Ionosphere model (IRI) was also released [*Rawer et al.*, 1978] ,this will be further explained in Subsection 4.1.1.

During the 1980s, two very important coupled models were released. In the following years, they both were subjected to continuous improvement and extensions. One of them is a first-principles model developed by the National Center for Atmospheric Research (NCAR) in the US. It began with the release of the Thermosphere General Circulation Model (TGCM) [*Dickinson et al.*, 1981]. It was subsequently extended to include the model for the coupled thermosphere and ionosphere system, thus forming Thermosphere-Ionosphere General Circulation Model (TIGCM) [*Roble et al.*, 1988]. TIGCM will be later used as a basis for the Thermosphere-Ionosphere Nested Grid model (TING) [*W Wang*, 1998]. By including a solution for the low-latitude electric field, the TIGCM evolved to give TIE-GCM: Thermosphere-Ionosphere-Electrodynamics General Circulation Model [*Richmond et al.*, 1992]. A further improvement was done by incorporating the mesospheric region; the model was then called Thermosphere-Ionosphere-Mesosphere-Electrodynamics General Circulation Model (TIME-GCM) [*Roble and Ridley*, 1994].

Another extension came by coupling the TIE-GCM with dynamics of the magnetosphere modelled by the Ionosphere-Magnetosphere Model (IMM) of *Peymirat and Fontaine* [1994] from the Centre Universitarie of Velizy in France. The result was called the Magnetosphere-Thermosphere-Ionosphere Electrodynamics General Circulation Model (MTIE-GCM) [*Peymirat et al.*, 1998]. TIE-GCM is also the basis of a more recent coupled model developed by the University Corporation for Atmospheric Research (UCAR) called Coupled Magnetosphere-Ionosphere-Thermosphere (CMIT) [*W Wang et al.*, 2008].

The second series of coupled models started at University College of London (UCL). The combination of the aforementioned thermosphere model by *Fuller-Rowell and Rees* [1980] and a convection model by *Quegan et al.* [1982] permitted the creation of the Coupled Thermosphere-Ionosphere Model (CTIM) [*Fuller-Rowell et al.*, 1996]. This was then extended in the Coupled Thermosphere-Ionosphere-Plasmasphere model (CTIP) [*Millward et al.*, 1996], and later in the Coupled Thermosphere-Ionosphere-Plasmasphere Electrodynamics model (CTIPe) [*Millward et al.*, 2001]. Another extension of the UCL CTIP called the Coupled Middle Atmosphere and Thermosphere (CMAT) General Circulation Model was proposed [*Harris*, 2000; *Harris et al.*, 2002].

It evolved into Coupled Middle Atmosphere and Thermosphere-2 (CMAT2) GCM thanks to *Yigit et al.* [2009].

In the late 1980's, a further coupled model, the Upper Atmosphere Model (UAM), was developed in the Moscow State Technical University (MSTU) [*Namgaladze et al.*, 1988].

Later during the 1990s, the Parameterized Ionospheric Model (PIM) [*Daniell Jr et al.*, 1995] was released. PIM is the combination of three models, GTIM, TDIM, and an empirical model for the plasmasphere. PIM was also integrated into the Parameterized Real-Time Ionospheric Specification Model (PRISM) [*Daniell Jr*, 1991; *Daniell Jr and Brown*, 1995], which is likely the first example of data assimilation applied to forecasting the ionospheric medium.

In 1995 The Field Line Interhemispheric Plasma Model (FLIP) by *Richards et al.* [1995] was released. This is a first-principles, mono-dimensional, time-dependent ionospheric model of ionosphere plasmasphere and thermosphere. Its algorithms provide solutions along entire magnetic flux tubes; the combination of more tubes produces a three-dimensional reconstruction.

The beginning of the new millennium witnessed the creation of another series of physics-based models. *Huba et al.* [2000] released the Sami2 is Another Model of the Ionosphere (SAMI2). It is a first-principles low latitude ionospheric model developed by the US Naval Research Laboratory (NRL) in Washington DC. SAMI3 is Also a Model of the Ionosphere is an extension of SAMI2 with a wider latitude range. Meanwhile in Australia, the Global Plasmasphere Ionosphere Density (GPID) [*Webb and Essex*, 2001] was developed in La Trobe University. One year later, *Ridley and Liemohn* [2002], from the University of Michigan, presented the Ridley Ionosphere Model (RIM). This was then used in the coupled model Global Ionosphere Thermosphere Model (GITM) [*Ridley et al.*, 2006]. In 2004, the Utah State University completed the Ionosphere-Plasmasphere Model (IPM), which was developed specifically for DA purposes [*Schunk et al.*, 2004]. In particular this is implemented in the Global Assimilation of Ionospheric Measurements (GAIM) that will be better explained in Section 4.2.2. A new coupled model was recently developed in the National Institute of Information and Communications Technology (NICT) of Japan. The new-born is called Ground-to-topside model of Atmosphere and Ionosphere for Aeronomy (GAIA) [*Jin et al.*, 2011].

4.1.1 Empirical models

In the late 1960s the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI) joined together to sponsor the International Reference Ionosphere (IRI) project. The aim was to provide an international standard for the specification of ionospheric parameters. In practice this evolved into the development of an empirical ionospheric model called IRI model [*Bilitza*, 1990] by using historical collection of all available ionospheric data coming from different sources. There have been a series of model versions, the latest is the IRI-2012 and

is able to provide, for a given location and time, monthly averages of electron density, electron and ion temperature, ion composition and many other parameters. The range within IRI-2012 works goes from 60 km to 2000 km above sea level. IRI is continuously improved and its data-base updated by an international group of experts.

The NeQuick model was developed from a profiler model by *Di Giovanni and Radicella* [1990] at the Aeronomy and Radiopropagation Laboratory (now T/ICT4D Laboratory) of the Abdus Salam International Centre for Theoretical Physics (ICTP) in Trieste (Italy) with the collaboration of the Institute for Geophysics, Astrophysics and Meteorology of the University of Graz (Austria). Mainly used for trans-ionospheric propagation application, NeQuick provides a fast empirical model of the ionosphere. The latest version, NeQuick2 by *Nava et al.* [2008] is used as default option in the IRI model to generate the top-side of the ionosphere.

The Mass Spectrometer Incoherent Scatter (MSIS) model is an extremely well known and largely accepted empirical tool for the estimation of temperatures and densities of the neutral species in the ionosphere. It was first presented with the name MSIS-86 by *Hedin* [1987] as the upper portion of the CIRA-86. CIRA stands for 'COSPAR (Committee on Space Research) International Reference Atmosphere' and it is an empirical model of the whole atmosphere. The MSIS model evolved to updated versions like the US Naval Research Laboratory NRLMSISE-00 [*Picone et al.*, 2002]. This version takes into account solar activity drivers (*F10.7* index) and geomagnetic perturbation (A_p index).

Another empirical model was released with the first version of MSIS (1987): the Horizontal Wind Model (HWM) by *Hedin et al.* [1988]. Also this one is an established model for the simulation of the horizontal wind in the upper atmosphere and its latest version is the HWM07 [*Drob et al.*, 2008].

4.1.2 Schematic classification of the major models

Section 4.1 reviewed existing ionospheric model in chronologic order. This subsection collates and categorizes some of them according to the criteria mentioned in the Introduction of this chapter. Table 1 is inspired by the Guide to Reference and Standard Ionosphere models [*AIAA*, 1999]. It has been modified and corrected according to updated information.

Model	Type	Geographic region	Altitude range (km)	Output	Species considered	Application
USU/TDIM	Physics-based	Global	90-1000	N_e, N_i, T_e, T_i	$NO^+, O_2{}^+, N_2{}^+, O^+, N^+, He^+$	Scientific studies
NCAR/ TIE-GCM	Physics-based	Global	97-500/700 According to solar activity	N_e, N_i, N_n T_e, T_i, T_n, neutral wind, etc.	$NO^+, O_2{}^+, N_2{}^+, O^+, N^+, O, O_2, NO, N(^4S), N(^2D)$	Scientific studies
CTIM	Physics-based Coupled	Global	80-10000	N_e, N_i, N_n, T_e, T_i, T_n	$NO^+, O_2{}^+, N_2{}^+, O^+, N^+, H^+, O, O_2, N_2 NO, N(^4S), N(^2D)$	Scientific studies
IRI	Empirical	Global	Ne: 50-2000 T: 120-3000 Ni:100-2000	N_e, N_i, T_e, T_i	$NO^+, O_2{}^+, N_2{}^+, O^+, N^+, H^+, He^+$	Various
SUPIM	Physics-based	Global	90-22000	N_e, N_i, T_e, T_i	$NO^+, O_2{}^+, N_2{}^+, O^+, H^+, He^+$	Theoretical studies of climatology
FLIP	Physics-based	Mid-latitudes	90-22000	N_e, N_i, T_e, T_i	$NO^+, O_2{}^+, N_2{}^+, O^+, He^+, O^+(^2D), O^+(^2P), N(^2D), N(^2D), NO, O(^1D), N_2(vib)$	Mid-latitudes studies
IFM	Physics-based	Global	90-15000	N_e, N_i, T_e, T_i, plasma drift velocities	$NO^+, O_2{}^+, N_2{}^+, O^+, N_2^+$	Scientific studies (Forecasting)
IPM	Physics-based	Global	90-30000	N_e, N_i, T_e, T_i	$NO^+, O_2{}^+, N_2{}^+, O^+, H^+, He^+$	Scientific studies
SAMI3	Physics-based	Global	90-20000	N_e, N_i, T_e, T_i	$NO^+, O_2{}^+, N_2{}^+, O^+, H^+, He^+, N^+$	Scientific studies
MSIS	Empirical	Global	0-1000	N_n, T_n	O, O_2, N, N_2, H, A, He	Various

Table 1. Schematic summarization of the major ionospheric models extrapolated from the Guide to Reference and Standard Ionosphere Models [*AIAA*, 1999] and adjusted.

4.2 Ionospheric data assimilation schemes

4.2.1 IDA3D (and IDA4D)

The Ionospheric Data Assimilation Three Dimensional (IDA3D) [*Bust et al.*, 2004] was developed at the Applied Research Laboratories of the University of Texas at Austin (US). IDA3D is based on a three-dimensional variation analysis (3DVar) and it is able to handle different types of ionospheric data in terms of electron density or electron content. The data sources for the used observation are: GPS ground-receivers, GPS occultation receivers, satellites for in situ measurements and beacon arrays. The model adopted in the scheme is normally IRI or the first-principles TIME-GCM. A Gauss-Markov Kalman filter is used to update the analysis and the error

covariance matrices. The output of IDA3D are the spatial analysis of electron density values for given times. IDA3D have recently evolved into The Ionospheric Data Assimilation Four Dimensional (IDA4D) [*Bust et al.*, 2007].

4.2.2 USU GAIM

The Global Assimilation of Ionospheric Measurements (GAIM) [*Schunk et al.*, 2005a; b; *Schunk et al.*, 2004] was developed at Utah State University under the Multidisciplinary University Research Initiatives (MURI) sponsored by the US DoD. GAIM uses physics-based ionospheric-plasmaspheric models combined with a Gauss-Markov Kalman Filter. It is able to handle various ionospheric measurements in real time (or near real time) from both ground-based and space-based platforms. The data sources used in this approach can include ionosonde, GPS ground-receivers, GPS satellite-receiver for occultation, DMSP (Defense Meteorological Satellite Program) for electron density in-situ observations, satellite-based instruments for line-of-sight Ultra Violet (UV) radiances, radio beacon and magnetometers. The principal physics based model adopted is the IFM, however IPM is also used in some USU-GAIM versions. One of the latest versions utilizes a physics-based reduced-state Kalman filter [*Scherliess et al.*, 2004]. The main output of the USU-GAIM is a three-dimensional time-dependent global characterization of the ionosphere in terms of electron density. In addition to this, it can provide different ionospheric drivers such as neutral winds and densities, magnetospheric and equatorial electric fields, and electron precipitation patterns. At the moment the USU-GAIM is operational at the Air Force Weather Agency (AFWA) and the NASA (National Aeronautics Space Administration) Community Coordinated Modelling Center (CCMC).

4.2.3 JPL/USC GAIM

The Global Assimilative Ionospheric Model (GAIM) [*Hajj et al.*, 2004; *Pi et al.*, 2003; *C Wang et al.*, 2004] was developed by the Jet Propulsion Laboratory and University of Southern California, leaders of the second consortium of the MURI program. The development of this scheme started in the 1999 and its purpose is to provide a system to monitor and forecast the ionosphere. JPL/USC GAIM includes a first principle ionospheric model, a supplementary model to generate driving forces, a data processing module and an optimization subsystem. The optimization is implemented by a Kalman Filter and 4-dimensional variational analysis (4DVar) approaches. The Kalman Filter performs covariance estimations and state correction from the ingested measurements and provides them to the model. The 4DVar approach estimates model parameters from the measurements and feeds the model with the adjusted drivers. The data sources that JPL/USC GAIM can consider are numerous and they include: ionosondes, ground GPS receivers, spaceborne GPS receivers (e.g. COSMIC), different satellite sensors such as the Limb Extreme UV (EUV) sensors (e.g. in Special Sensor Ultraviolet Limb Imager (SSULI) and Special Sensor Ultraviolet Spectrographic Imager (SSUSI)), and radio beacons.

4.2.4 EDAM

The Electron Density Assimilative Model (EDAM) [*Angling and Cannon*, 2004; *Angling and Khattatov*, 2006] was developed at QinetQ and sponsored by the UK Ministry of Defense. The data sources of EDAM are satellite (for radio occultation) and ground based GPS receivers. PIM is adopted as background model, and its data ingestion approach is an application of the Best Linear Unbiased Estimator (BLUE).

4.2.5 MIDAS

The Multi Instrument Data Analysis System (MIDAS) [*Mitchell and Spencer*, 2003; *Spencer and Mitchell*, 2007] was developed at the University of Bath. The current version of MIDAS is correctly defined as a tomographic imaging approach; however it is included in this section because it can be viewed as a four dimensional Data Assimilation scheme without a background model. In one of its earlier versions [*Spencer and Mitchell*, 2007], MIDAS was actually used as a genuine assimilative method by implementing a Kalman filter and including a physics based model. This set up was successfully used for the detection of small-scale, fast-moving structures in Polar Regions.

The latest version of MIDAS provides three-dimensional reconstructions of the ionosphere in terms of electron density. The observations used are STEC collected by GPS ground-receivers. MIDAS implements a Tikhonov regularization (Eq. (7.3)) where a priori information (normally second derivative matrix) is applied in order to obtain a smoother solution. Strong constraints of the vertical profile are used in the inversion, in the form of orthonormal basis functions. They are generated by applying Singular Value Decomposition (DSV) methods to a priori information provided by IRI-95 or an ensemble of predefined Chapman profiles. In the first version of MIDAS [*Mitchell and Spencer*, 2003], horizontal basis functions were also included.

Details about MIDAS algorithms and examples of its output can be found in Section 7.1 and in Appendix A respectively. This dissertation will discuss the methodology with which MIDAS can be upgraded to a full-physics data assimilation scheme with forecasting capabilities.

Summary

This chapter presented the most relevant existing ionospheric models. In addition, the principal examples of ionospheric DA were described.

The next Chapter introduces A New Ionospheric Model (ANIMo) and explains its algorithms. According to the classifications criteria that have been defined in the Introduction of this chapter, it is possible to categorize ANIMo as physics-based model. It was designed to be applied at mid-latitude regions in an altitude range that goes from 80 to 600 km. For a given time and location, ANIMo provides outputs in terms of electron and ion densities. The considered species are NO^+,

O_2^+ and O^+. The applications of ANIMo are: supporting the ionospheric tomography imaging, DA and forecasting. Similarly with what has been done for the major ionospheric models in 4.1.2, it is possible to translate the specifics of ANIMo in a more schematic manner (Table 2).

Model	**Type**	**Geographic region**	**Altitude range (km)**	**Output**	**Species considered**	**Application**
ANIMo	Physics-based	Mid-latitudes	80-600	N_e, N_i	NO^+, O_2^+, O^+	Supporting ionospheric imaging, now-casting, forecasting

Table 2. Schematic description of the characteristics of ANIMo.

Chapter 5 A New Ionospheric Model (ANIMo)

Abstract

A New Ionospheric Model (ANIMo) is a physics-based ionospheric model specifically built for supporting tomographic imaging and Data Assimilation of the ionised upper atmosphere. For any given point in the ionosphere, the model simulates the primary sources of ions and electrons from the intensity of the Sun and the density of neutral particles. The intensity depends on the geometry of incident radiation (for both diurnal and seasonal variations), the absorption due to neutral densities and on the solar activity. The considered neutral species are N_2, O_2, and O, which produce the ions NO^+, O_2^+, and O^+ through photoionization and related processes. The model approximates the electron density as a sum of these ions densities. The recombination (loss term) is integrated in the calculation by using recombination rates that depend on the charged particle density and temperature. Regarding the transport processes, the model considers the vertical motion due to the ambipolar plasma diffusion. The vertical diffusion velocity is adjusted to simulate the daily descending drift due to ionospheric wind. A downward flux of ion is added as top-side boundary condition.

Introduction

The objective of the overall project was to develop an ionospheric model. There are many approaches to the formation of an ionospheric model however the choice of approach is highly constrained by the intended applications. Firstly, the amount of effort, both in preparing the model and in running it (e.g. computational requirements), must be appropriate with respect to the level of accuracy the model needs to achieve. For this project, we concluded that the model should contain enough physics to enable it to be accurate over mid-latitude regions where the GPS data coverage can be sparse. Further, that the physics would allow the model to be used for short-term forecasting of the ionosphere, again over mid-latitude regions.

Modelling the ionosphere is a very challenging task due to the complexity of the ionosphere and the number of involved dynamics and variables. A further complication is that the ionosphere is a coupled system. The model requires a number of forcing parameters to be defined in order to have an accurate solution. These forcing parameters are, for example, the density of neutral species, the ion and electron temperatures and the strength of the neutral wind. Unfortunately, the appropriate ionospheric measurements to provide these parameters are often not available and therefore it is difficult to estimate the best set-up. A further complication is not knowing the sensitivity to the

effects of tuning these forcing parameters. It can be very challenging to fully understand the complex behaviour of the model because of the total interactions of its parts.

The decision of developing ANIMo for supporting ionospheric tomography was essentially driven by the requirements of a DA system. The model characteristics are described in Section 5.1 of this chapter. Section 5.2 is dedicated to the description of ANIMo and the technical solutions that have been applied. It is divided in three subsections: the first one (5.2.1) regards the aeronomy portion of the model, the second (5.2.2) concentrates on the transportation processes and the last (5.2.3) is addressed to the model continuity equation and the numerical technique used to find the final solution. Consistently with the rest of the results chapter, the final section is a summary.

Some of the material in this chapter has been published [*Da Dalt et al.*, 2014].

5.1 ANIMo requirements

ANIMo is a physics-based ionospheric model built into computer software using MATLAB. The advantages of using a first-principle model are many. Firstly, it is preferable to avoid using empirical models in DA approaches, especially when forecasting, because climatological models are poor at forecasting specific events. Secondly, the usage of a physical model allows more flexibility in the manner the model is used. This includes, for example, the possibility of simulating specific unsettled conditions, studying their evolution and using the model itself in interpreting unclear ionospheric dynamics and their connections. The model considers those principal ionospheric processes that can guarantee a reasonable level of accuracy. Complex processes such as dynamical effects like tides, travelling ionospheric disturbances, joule heating and electric fields are not simulated in the model however they may be the subject of future research and development.

The model also needs to be robust and stable. These characteristics are usually pursued when developing any sort of mathematical model. In this case, the possibility of simulating extreme conditions makes these requirements more important to keep the model reliable. The final feature refers to the computing effort. The model was coded by maintaining a certain efficiency, this is not only important for the developer but also for its future implementation, possibly in an operative system used for now-casting and forecasting applications.

5.2 ANIMo description

ANIMo performs on three dimensional grids defined by latitude, longitude and altitude and it is designed for the mid-latitude regions in an altitude range from 80 to 600 km above sea level. Its input parameters, in addition to location and time, are: A_p index and *F10.7* (which are retrieved automatically from online repositories). There is also the capability to allow adjustment of the

vertical velocity and the ion downward flow value (topside boundary condition) which can be defined by the user at the beginning of each run. The outputs of ANIMo are three-dimensional simulations of the ionosphere; it is able to provide electron density values and the principal ion species O^+, O_2^+, and NO^+. To better explain the functioning of the model, ANIMo is described here in three stages. The first one is focussed on the chemistry of generation and recombination processes. The second is more physically oriented, where the considered transportation processes are explained. The third describes the applied numerical approach. The closest existing model to ANIMo is the FLIP model by *Richards et al.* [1995]. The similarity to ANIMo is found mainly in the modelling of the chemical dynamics.

5.2.1 Aeronomy: generating reaction rates

ANIMo is a reformulation and expansion of the model of generation and recombination from the initial work of *Benton and Mitchell* [2012] with which it shares mainly the aeronomic calculations. Starting from the generation part, ANIMo uses the EUV flux model for Aeronomic Calculations (EUVAC) by *Richards et al.* [1994b] for modelling photoionization processes. EUVAC provides solar fluxes values divided in 37 wavelength bins. These are based upon the F4113 solar reference flux spectrum measured in April 1974. The bins overall values range from 50 Å to 1050 Å, therefore wavelength from X-rays to extreme ultraviolet are considered. Some of the bins collect wavelength ranges, always 50 Å wide, other refer to specific emission lines. In addition to the reference spectrum, the EUVAC model provides solar activity proxies and, for each bin, a solar activity scaling factor. These are necessary to calculate flux values according to the solar activity situation that has to be defined by setting *F10.7* and *F10.7*A[1] parameters.

In order to calculate the photoionization rates, ANIMo simulates the flux attenuation due to the solar absorption by the most abundant neutral species in the ionosphere: O, O_2 and N_2. The MSIS model is used to simulate the densities of the neutral species; in particular one of its latest versions, NRLMSISE-00, is adopted [*Picone et al.*, 2002]. The MSIS model is susceptible to the solar activity status and the geomagnetic conditions. For this reason, it requires to be fed with *F10.7* and *F10.7*A parameters and the definition of several A_p indices, partly retrieved from previous times. The absorption cross sections of O_2 and N_2 are taken from the EUVAC model (Table 3 of the work of *Richards et al.* [1994b] and reproduced here in Table 3), the values for O are assumed equal to its ionization cross sections (reported in Table 2 in the paper of *Richards et al.* [1994b] and reproduced here in Table 3) apart from the 29th and 31th wavelength bins of the EUVAC parameterization of the solar spectrum. The absorption cross sections of the atomic oxygen for these two bins were recalculated on the basis of Table B from the work of *Fennelly and Torr* [1992] that reports a series of wavelengths and relative values of absorption and ionization cross sections which are not

[1] *F10.7*A is the 81 days, 40 before and after the selected date, average of the daily *F10.7* parameter.

considered identical. An adaptation of these values to the EUVAC bins arrangement is necessary. The absorption cross section of the 36th bin for bi-atomic oxygen is amended according to the EUVAC corrections paper by *Richards et al.* [1994a].

Wavelength bin number	O_2	N_2	O
1	1.316	0.72	0.73
2	3.806	2.261	1.839
3	7.509	4.958	3.732
4	10.9	8.392	5.202
5	13.37	10.21	6.05
6	15.79	10.9	7.08
7	14.387	10.493	6.461
8	16.8	11.67	7.68
9	16.81	11.7	7.7
10	17.438	13.857	8.693
11	18.32	16.91	9.84
12	18.118	16.395	9.687
13	20.31	21.675	11.496
14	21.91	23.16	11.93
15	23.101	23.471	12.127
16	24.606	24.501	12.059
17	26.04	24.13	12.59
18	22.72	22.4	13.09
19	26.61	22.787	13.024
20	28.07	22.79	13.4
21	32.06	23.37	13.4
22	26.017	23.339	13.365
23	21.919	31.755	17.245
24	27.44	26.54	11.46
25	28.535	24.662	10.736
26	20.8	120.49	4
27	18.91	14.18	3.89
28	26.668	16.487	3.749
29	22.145	33.578	6.545
30	16.631	16.992	3.498
31	8.562	20.249	6.3104
32	12.817	9.68	1.315
33	18.73	2.24	0
34	21.108	50.988	0
35	1.63	0	0
36	1.05	0	0
37	1.346	0	0

Table 3. Absorption data (and parameterization of UV spectrum) used in ANIMo to simulate the attenuation of the incident solar ray.

For a selected point in the ionosphere, the solar attenuation is therefore simulated by considering the atmospheric segment of the incident solar ray-path; this is set to start at 1000 km altitude. This was chosen because it is necessary to consider the presence of neutral species responsible for the solar attenuation also over 600 km. This means that, despite the fact that 600 km is the upper limit of ANIMo results, the part of the model responsible for retrieving neutral compositions (MSIS model) works also over that limit up to 1000 km. For a number of points along the aforementioned segment, one every 5km, a partial attenuated spectrum is calculated. The final attenuation is built by incrementing its values until the end of the considered solar ray-path which corresponds to the aforementioned selected point. Here, to simulate the principal photoionization processes, the generated attenuated flux together with ionization cross sections and the respective neutral densities, is used to calculate the photoionization rates of the following reactions.

$$O \rightarrow O^+ + e^- \tag{5.1}$$

$$O_2 \rightarrow O_2^+ + e^- \tag{5.2}$$

$$N_2 \rightarrow N_2^+ + e^- \tag{5.3}$$

The reactions correspond to equations (2.1), (2.2), and (2.3) from Chapter 2 dedicated to the description of the ionosphere. The ionization cross sections are taken from Tables 2 and 3 of the EUVAC model [*Richards et al.*, 1994b] and presented here in Table 4, where the value of the 35th bin for O_2 is corrected [*Richards et al.*, 1994a].

Wavelength bin number	O_2	N_2	O
1	1.316	0.72	0.73
2	3.806	2.261	1.839
3	7.509	4.958	3.732
4	10.9	8.392	5.202
5	13.37	10.21	6.05
6	15.79	10.9	7.08
7	14.387	10.493	6.461
8	16.8	11.67	7.68
9	16.81	11.7	7.7
10	17.438	13.857	8.693
11	18.32	16.91	9.84
12	18.118	16.395	9.687
13	20.31	21.675	11.496
14	21.91	23.16	11.93
15	23.101	23.471	12.127
16	24.606	24.501	12.059
17	26.04	24.13	12.59
18	22.72	22.4	13.09
19	26.61	22.787	13.024
20	26.39	22.79	13.4
21	31.1	23.37	13.4
22	24.937	23.339	13.365
23	21.306	29.235	17.245
24	23.75	25.48	11.46
25	23.805	15.06	10.736
26	11.72	65.8	4
27	8.47	8.5	3.89
28	10.191	8.86	3.749
29	10.597	14.274	5.091
30	6.413	0	3.498
31	5.494	0	4.554
32	9.374	0	1.315
33	15.54	0	0
34	13.94	0	0
35	1.05	0	0
36	0	0	0
37	0.259	0	0

Table 4. Ionization data (and format of UV spectrum) used in ANIMo to estimate photoionization rates to simulate ionization processes.

Neutral densities are supplied by the MSIS model. The rate of the last reaction is in reality estimated to calculate the production of the ion species NO^+ by assuming that Equation (5.3) is followed by a very rapid reaction:

$$N_2^+ + O \rightarrow N + NO^+ \tag{5.4}$$

Regarding the recombination chemistry for the monoatomic oxygen ion, ANIMo considers two ion rearrangements reactions:

$$O^+ + O_2 \xrightarrow{k_R1} O + O_2^+ \tag{5.5}$$

$$O^+ + N_2 \xrightarrow{k_R2} N + NO^+ \tag{5.6}$$

In each reaction, the k_R parameter refers to the recombination rate coefficients. It is possible to notice that the direct recombination of O^+, described in Eq. (2.19), is here not contemplated because it is quantitatively insignificant [*Benton and Mitchell*, 2012]. The reported reactions Equations (5.5) and (5.6) are also responsible for the production of the secondary ions O_2^+ and NO^+. For these two species, ANIMo takes into account their dissociative recombination (Eq. (2.22)and (2.20)) reactions.

$$NO^+ + e^- \xrightarrow{k_R3} N + O \tag{5.7}$$

$$O_2^+ + e^- \xrightarrow{k_R4} O + O \tag{5.8}$$

The recombination coefficients k_R1, k_R2, k_R3, and k_R4 are taken from the work of *Torr and Torr* [1979]. They depend on ion temperature which is modelled by IRI-2012.

5.2.2 Transportation mechanisms

The transportation processes considered by ANIMo refer mainly to the vertical diffusion of the monoatomic oxygen ion. A basic ambipolar diffusion equation taken from the work of *Rishbeth and Garriott* [1969] is used to generate the vertical movement of O^+. The original formula (Eq. (2.26)) is applied in ANIMo in a slightly modified form:

$$Wd = -sinI^2 \left(\frac{k_B\, 2T_p}{m_{O^+}} \frac{d[O^+]}{dz} \frac{1}{[O^+]} + g \right) \frac{1}{\nu_{O^+\text{-}O}} \tag{5.9}$$

Where Wd is intended to be the diffusion vertical velocity for the ion species O^+, in units $(m\ s^{-1})$. I is the geomagnetic dip or inclination, which is taken from the IGRF, in particular from the year 2000 version [*Finlay et al.*, 2010]. The introduction of the dip angle differs from the implementation of the original formula. It takes into account the movement constraint due to the presence of the geomagnetic field – effectively this will slow the diffusion where the magnetic field is not vertical. This method has been already used in other ambipolar diffusion equations, for example by *Salah and Holt* [1974]. k_B is the Boltzmann constant $(kg\ m^2\ s^{-2}\ K^{-1})$ while T_p, in (K), is a parameter called plasma temperature and it is calculated by dividing the sum of ion and electron temperatures by two [*Rishbeth and Garriott*, 1969]. In ANIMo the necessary temperature values are provided by IRI-2012. The element $d[O^+]/dz$ is the derivative of the O^+ ion density (indicated with square brackets) along the altitude z, its unit can be considered $(n\ m^{-3})$. m_{O^+} is the mass (kg) of O^+ and

g is the gravitational acceleration ($m\ s^{-2}$). ANIMo takes into account that the value of the latter decreases with the altitude. In fact this is significant as the value ranges from 9.6785 m/s at 80 km to 8.9565 m/s at 600 km. $\nu_{O^+\text{-}O}$ stands for ion-neutral atomic oxygen collision frequency and it is calculated using the following formula by *Salah* [1993]:

$$\nu_{O^+\text{-}O} = 4E^{-17}\ T_r^{0.5}\ [O] \tag{5.10}$$

The collision frequency parameter, having the units of a traditional frequency (s^{-1}), depends on temperature and on the amount of neutral oxygen present $[O]$. The temperature value used, T_r, is the half of the sum of the ion and neutral temperatures.

During daytime, the vertical velocity is adjusted by a correction factor that changes with altitude. This factor is one of the input parameters of the model and it is introduced to simulate the additional daytime downward drift component due to the ionospheric wind system [*Kohl and King*, 1967]. The function with which this correction varies along the altitude follows a Gaussian distribution centred in the highest altitude point reached by the model.

The definition of the shape and velocity correction value is not totally arbitrary. During the validation phases of ANIMo, it was necessary to 'curb the enthusiasm' of the modelled photoionization processes during daytime. The reason of the particular shape is related to the acceleration of the daily downward drift to lower the altitude of the electron density peak. At the same time, this function avoided upsetting the model dynamics for the bottom-side of the vertical profile. This adjustment is considered an acceptable arrangement to compensate for the fact that otherwise the plasma movement would be governed solely by the vertical diffusion. Extensive private communication with Jan J. Sojka from Utah State University [*Sojka et al.*, 2013] led to the decision to introduce this modification.

Adjusted velocity profiles produced by ANIMo have been compared to ISR measurements of vertical plasma drift velocities from the validation case studies to concur that this was a sensible approach. The correction is activated by an enhancing photoionization rate at certain altitudes. When the photoionization rate is diminishing, the correction is disabled. The night-time maintenance is therefore ensured by the continuous downward ion flux that will be described in the next subsection.

5.2.3 Continuity equations and numerical solving approach

ANIMo is based on the solution of the continuity equation for the monoatomic oxygen ion for a given vertical profile:

$$\frac{\partial[O^+]}{\partial t} = Q_{O^+} - L_{O^+}([O^+]) - \frac{\partial([O^+]v)}{\partial z} \tag{5.11}$$

Where, for a given point in the ionosphere, $\partial[O^+]/\partial t$ is the rate of change of O^+ density over time t and it is assumed as the combination of production Q_{O^+}, loss L_{O^+} and vertical transportation $\partial([O^+]v)/\partial z$ processes. The production term Q_{O^+} refers to the O^+ photoionization rate of the reaction Equation (5.1), estimated by following the procedure explained in the previous subsection. The loss term is related to the recombination rates of the ion rearrangement reactions Eq. (5.5) and (5.6). The value of this term is therefore found by:

$$L_{O^+} = k_R1([O_2]) + k_R2([N_2]) \tag{5.12}$$

The aeronomy section (5.2.1) introduced the coefficients k_R1 and k_R2, the neutral densities values are, once again, estimated by the MSIS model. The transportation term $\partial([O^+]v)/\partial z$ considers mainly the vertical ambipolar diffusion. v is in fact a generic velocity value, along the altitude z, that in ANIMo corresponds to the aforementioned (Section 5.2.2) adjusted vertical velocity.

To solve numerically the continuity equation for the monoatomic oxygen ion, ANIMo utilizes an explicit method. Also known as the forward Euler method, this scheme belongs to the finite difference methods. These approaches are fundamental in computer simulations of physical processes such as continuity equations. They solve differential equations by an approximation of the derivative by finite difference equations and they require a discretization of time (in time steps) and space (in cells, pixels or voxels depending on the number of the considered dimensions). From a time point of view, explicit methods approximate the derivative performing a forward difference in time to find the solution of future time steps. The name comes from the fact that this solution is found explicitly from the known values at the current time. These approaches can be applied for solving partial differential equations, which can be used to describe diffusive phenomena. In this case the method requires a discretization not only in time but also in space.

Continuing with the diffusion example, in order to model the evolution of a certain diffusive property in a discrete location (cell), it is necessary to consider the contribution from and to the contours of the system (on a basic mono-dimensional case, the contours are the cell in front and the one behind the selected location). In this case, by 'contours' what is meant is the pixel (cell) above and below the one at a given altitude. The evolution of the central cell can be modelled by approximating the derivative forward and backward in space – the scheme is therefore known as Forward Time Centered Space (FTCS). This type of scheme is used by ANIMo in the solution of the monoatomic oxygen ion continuity equation. The space domain in the model corresponds to a vertical profile discretized in several height steps. For each time step t and each profile cell i the solution is solved by:

$$[O^+]_i^t = [O^+]_i^{t\text{-}1} + {Q_{O^+}}_i^{t\text{-}1}\Delta t - {L_{O^+}}_i^{t\text{-}1}\left([O^+]_i^{t\text{-}1}\right)\Delta t + \frac{[O^+]_{i+1}^{t\text{-}1} v_{i+1}^{t\text{-}1} - [O^+]_{i\text{-}1}^{t\text{-}1} v_{i\text{-}1}^{t\text{-}1}}{\Delta t \Delta z} \quad (5.13)$$

In ANIMo the explicit scheme is applied to find the current solution $[O^+]_i^t$ from the previous state of the system, indicated on each element of the equation by t-1. The elements ${Q_{O^+}}_i^{t\text{-}1}$ and ${L_{O^+}}_i^{t\text{-}1}$ are therefore the previous production and loss for recombination terms respectively. Δt is the adopted time interval; a brief discussion regarding its definition will be reported later in this paragraph. The indented element on the second line of the equation is responsible for estimating the contributions due to transportation processes. $[O^+]_{i+1}^{t\text{-}1}$ and $[O^+]_{i\text{-}1}^{t\text{-}1}$ are the monoatomic ion density of the adjacent cells, respectively the one on top and the one underneath the current cell.

Similarly, $v_{i+1}^{t\text{-}1}$ and $v_{i\text{-}1}^{t\text{-}1}$ are the velocity associated with the neighbour cells. As explained, v corresponds to an adjusted diffusion velocity calculated specifically for each position. Δz is the space interval, which in this case is the height of one of the cells. The definition of Δt and Δz is crucial when using numerical methods because it can affect the numerical stability of the mathematical analysis. The stability attribute is associated with the accuracy of the approximation in solving the derivatives. If this attribute is not guaranteed, the method produces highly inaccurate solutions and can eventually crash the executing program. Explicit methods are conditionally stable, which means that the stability can be achieved and maintained under certain conditions that are related to the choice of Δt and Δz.

Normally, the definition of the intervals is done in relation to the value of the evolution term, in this case the velocity. For ANIMo this definition is very difficult to set because velocities vary at each time step and location. Therefore, through a process of experimentation, it was decided to use a fixed altitude interval of 10 km height and a time interval that diminishes at night. These interval values were set after running tests aimed to achieve stability. The values of the two intervals are related and their effect interchangeable, hence it is in theory possible to set bigger Δt and smaller Δz. As mentioned, the time interval is smaller during night. This approach is necessary to deal with extremely low ion densities, hence velocities, which occur during night-time. The change of Δt is triggered by a simple switch referred to Local Time (LT).

A similar stability issue happens at the bottom of the profile; below a certain altitude, O^+ densities are close to zero and this tends to compromise the stability. For this reason the presented numerical solution operates between fixed altitudes. Below this fixed heights range, O^+ densities are considered equal to zero therefore transportation processes are not applied. This enables more freedom for defining the time interval.

A further point of note is the definition of the boundary conditions. Regarding the bottom boundary condition, it was arbitrarily chosen that the ion density values are equal for the first two locations of the numerical solving profile. For the top boundary condition a downward ion flux is assumed and set constant during the model run. Its value can be defined as an external forcing parameter. It is applied in the numerical approach by multiplying it by $\Delta t/\Delta z$ and adding the result to the velocity of the highest cell. There is no relationship between the downward ion flux and the vertical velocity adjustment introduced in Section 5.2.2. They can be defined independently in order to calibrate the model to simulate specific conditions.

Before introducing how $[O_2^+]$ and $[NO^+]$ are solved and how they fit into the presented approach, it is worth to step back and have a look at the whole of ANIMo's functionality. The monoatomic continuity equation is solved for each time step and for each cell along the considered altitude. This generates an O^+ density profile from which a velocity profile is calculated. The density and velocities profiles are then used in the next time step as previous elements to solve the continuity equation for the current time along the same altitude. Although this operation is performed each time step, the generation of its coefficients, temperatures and neutral densities occurs less frequently. In particular, photoionization rates are generated every 60 seconds. Temperature from which depends the calculation of recombination rates and part of the diffusion velocities is estimated by IRI-2012 every 30 minutes. The MSIS model, for neutral densities for the calculation of absorptions, photoionization processes and velocities, performs every 180 minutes. The decision to perform these operation at different time steps with respect to the one used for the numerical method is completely sensible because it relates to the timescales over which the parameters change. It makes sense to think that the difference, for example between IRI-2012 temperatures estimations calculated less than 30 minutes apart, is not worth taking into account. Furthermore this reduces machine calculation effort and hence produces a more efficient code. The determination of the frequency for each of the mentioned operations was not random, it resulted from numerous tests specifically performed for this purpose.

Similarly, the calculation of the secondary ions densities, $[O_2^+]$ and $[NO^+]$, does not occur at each Δt. Because the chemistry of these species is much faster with respect to the chemistry of O^+, it is assumed that their densities and photoionization rates are in equilibrium with the current ones of O^+. This means that there is no need for numerical solving and therefore their estimation can be performed for the whole profile (80-600 km) only when requested. The following equations describe how $[O_2^+]$ and $[NO^+]$ are estimated:

$$[O_2^+] = k_R1[O^+] + \mathrm{Q}_{O_2^+} - k_R3[e^-][O_2^+] \tag{5.14}$$

$$[NO^+] = k_R2[O^+] + \mathrm{Q}_{NO^+} - k_R4[e^-][NO^+] \tag{5.15}$$

Where k_R1 and k_R2 are the coefficients of the O^+ recombination reactions shown in Equations (5.5),(5.6) and adopted in the loss Equation (5.12). $Q_{O_2^+}$ and Q_{NO^+} are photoionization rates for O_2^+ (Eq. (5.2)) and NO^+ (Eq. (5.3)) respectively. k_R3 and k_R4 are the O_2^+ and NO^+ recombination coefficients, and they refer to reactions shown in Equations (5.7) and (5.8). Details about these coefficients and rates can be found in the previous paragraph. In agreement with the definition of plasma and by assuming that the O^+ density is zero from 80 km to 120 km, ANIMo consider that the electron density $[e^-]$ of the selected vertical profile is:

$$[e^-] = [O^+] + [O_2^+] + [NO^+] \tag{5.16}$$

As aforementioned, the estimation of $[O_2^+]$ and $[NO^+]$, and therefore of $[e^-]$, are executed only when requested. This means that the frequency of these operations is related to the choice of the ANIMo user depending on when and how often the model outcomes are required. ANIMo has a centralized system in order to deal with all the different intervals, which can be set at the beginning of every run. Once electron and ion densities are calculated for the whole profile, the latter can be combined with other profiles in order to have simulated three-dimensional reconstructions. To do so it is necessary to define a latitude and longitude grid as an input parameter.

ANIMo can also be initialized by entering a set of values to start the run. By default, IRI-2012 is used to produce the initial vertical profile of O^+ density, otherwise it can be introduced as an input by the user. Similarly other parameters, such as temperature and vertical velocity profiles can be easily introduced by the user. The possibility of doing so is extremely important; ANIMo can be used as a testing ground to simulate specific conditions to see how they evolve and ultimately how to deal with them when using data ingestion techniques.

Summary

One of the objectives of this doctoral project was to develop an ionospheric model to use in support of ionospheric tomography and eventually to implement in a DA scheme. In contrast with the other ionospheric models, ANIMo was built by always having in mind these final purposes. This implies that specific model requirements were complied with and that, on the other hand, it was possible to tolerate several assumptions.

ANIMo was developed from the model of generation and recombination from the work of Benton [*Benton and Mitchell*, 2012] which describes the major photoionization and recombination processes in the ionosphere. The latter was upgraded by implementing the EUVAC model to permit to take into account the solar activity when performing the simulation of plasma production processes. For similar reasons, also the MSIS model, responsible for providing neutral densities values in the atmosphere, was introduced. This was relevant in acquiring accuracy in the simulation of absorption and recombination processes and in the estimation of photoionization rates. IRI-2012

was then included for providing the default initialization of the model and for generating temperature profiles for the calculation of chemical reactions rates. The introduction of the transportation processes involved the implementation of a finite difference numerical method. In particular ANIMo simulates plasma movements mainly by vertical ambipolar diffusion, which depends on the presence of plasma, the collision frequency with neutrals, temperature and the magnetic inclination of the geomagnetic field. For this, once again, MSIS and IRI-2012 models are asked to provide necessary parameters. Diffusion velocities are adjusted in order to consider the diurnal downward drift and a downward flux is added as topside boundary condition. The final solution is calculated by an explicit method which is run at very small time steps in order to guarantee its stability. ANIMo can be used in mid latitudes to produce three-dimensional reconstructions of electrons and ions (NO^+, O_2^+, and O^+) densities, its altitude range of action goes from 80 to 600 km. Apart from the generation of densities of neutral species and temperatures, ANIMo is a first-principles physics model. It is stable and robust and, because it is driven by the principal ionospheric processes, it is efficient and non-complex. These features are extremely important requirements for its implementation in tomographic imaging and DA approaches. Furthermore, the module structure and the possibility of playing with its forcing parameters make ANIMo a unique tool for testing and experiment new solutions for ionospheric now casting and forecasting applications.

Chapter 6 ANIMo outputs and validation

Abstract

The ionospheric realizations generated by ANIMo show that the model is capable of reproducing realistic vertical electron and ions density profiles. By analysing these realizations over time it is possible to demonstrate that the model can also simulate diurnal variations of the ionosphere. A validation assessment tested the accuracy of the model; the latter was compared with real observations and other ionospheric models over several days in different periods of the year. Validation results demonstrate that ANIMo produces realistic results within the range of other models. Reliability and robustness of ANIMo were also investigated in the testing phase by means of different sensitivity tests. One of them was focussed on using temperature as forcing parameter to evaluate the behaviour of the model and to analyse how it copes with simulated extreme conditions. The testing period triggered a phase of adjustments and further assessments of the model. The presented results declare ANIMo suitable to be implemented in support of ionospheric tomography and for experimenting new solutions for improving ionospheric specification and prediction.

Introduction

The final stages of the development of a model involve both validation and refinement. There are different types of test used to understand whether the model performs according to the requirements that developers planned to achieve and each type can assess a different aspect of its usage. One of them consists of simply checking if the model is working and eventually if it responds correctly to variations of its input parameters. This could be less trivial that it sounds as very often, especially in Numerical Weather Forecasting (NWF), models have non-linear behaviour. The sensitivity analysis implies an investigation of the error propagation and uncertainty and consists of analysing the variation of the model outcomes resulting from a variation of forcing parameters or initial conditions. The test outcome is normally the detection of errors or missing information and leads to a phase of correction and refinements.

The next stage is to validate the model against real data. If there is a strong divergence between measured data and model outcomes, then it is fundamental to understand the reason for this departure and how the model can be improved. From the results of the validation test, it is also possible to understand the limitations of the model. The choice of the validation sets and of the comparison criteria is hence very important. The chosen criteria are the electron density at the

peak of the ionospheric profile (NmF2) and the altitude of this peak (hmF2). The selection is determined by two factors. First, it is vital that the model performs well for the above terms in order to support ionospheric tomographic imaging. Secondly, this is good practice in ionospheric model comparisons [*Anderson et al.*, 1998].

Other type of test can be executed in order to determine if the model is efficient; this is not only related to how it is coded but also to its general design. The aim is to eliminate the usage of redundant parameters and/or sub functions. The choice of which tests have to be done and how well the model needs to perform when they are executed is related to the final usage of the model.

In this dissertation coding tests were performed continuously during code development. Here the focus is more on the validation against real observations. This chapter describes the tests that were performed on ANIMo to understand if it satisfies the requirements that were defined to achieve the general objective of the project. Although five cases, one for each season plus one in perturbed conditions, were analysed in the validation process, one in particular (winter case) is used in the other tests in the next chapters. This is mainly due to the big computing effort required for running simulations and reconstructions. Furthermore, by referring to a particular case a certain consistency is maintained throughout the reading of the dissertation.

The first section of this chapter (Section 6.1) shows some examples of typical outcomes of ANIMo. Section 6.2 describes how the model was validated against real ionospheric measurements and the empirical model IRI-2012. In particular, Section 6.2 focusses on unperturbed case studies and Subsection 6.2.1 on a case study in unsettled conditions. A temperature sensitivity test is also presented and can be found in Section 6.3. Section 6.4 and 6.5 display several reconstructions of ANIMo for the winter case study by setting different top boundary conditions and vertical velocity adjustments respectively. Also for this chapter, the final paragraph is a summary of the presented results.

The present chapter is partly based on the work of *Da Dalt et al.* [2014].

6.1 ANIMo outputs

Chapter 5 summarised the technical design of ANIMo. The model generates ions and electron density vertical profiles for a specific geographical location. Three-dimensional reconstructions are then provided by combining multiple resulting profiles over a selected region. This means that each profile is simulated independently from the others and represents the fundamental aspect of a model – there is no horizontal coupling.

Figure 6 shows an example of vertical profile produced by ANIMo. In this case, the model was run to simulate the ionosphere above the location of the Millstone Hill Haystack Observatory (Lat.

42.6°, Lon. 288.5°) from an altitude of 80 to 600 km in 10 km steps. Regarding the date, the simulation refers to the 29th of December 2011 at 1400 LT (1900 UT), which was a quiet day from a space weather point of view (A_p 9; *F10.7* 142.3). In this particular case the upper boundary conditions and the velocity adjustments were set to the default values.

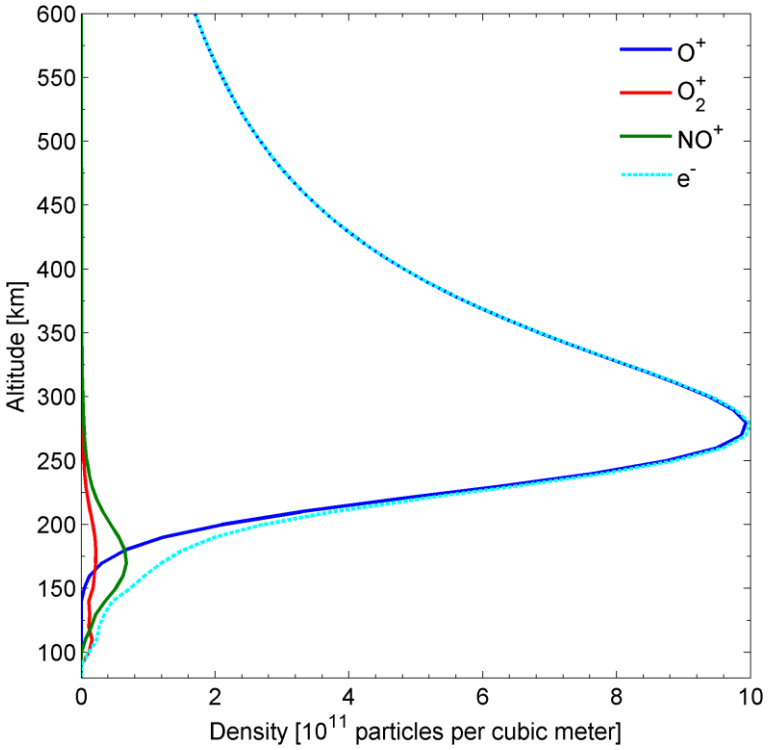

Figure 6. Ions and electron density vertical profiles produce by ANIMo above the Millstone Hill location (Lat. 42.6°, Lon. 288.5°) from 80 to 600 km on the 29th December 2011 at 1400LT (1900) UT (A_p 9; *F10.7* 142.3).

The result shown in Figure 6 is taken from a 2 days simulation run exhibited in Figure 7, in particular from the 29th to the 30th of December 2011 (A_p 9-7; *F10.7* 142.3-136.4). ANIMo was initialized at 0300 LT (0800 UT) by an O^+ vertical profile produced by IRI-2012 and let run for 48 hours in which an outcome in terms of electron density was saved every 30 minutes. Downward flux as boundary condition and velocity adjustment are the same used in the aforementioned single profile run.

The solid blue line refers to the density profile of the mono atomic oxygen ion (O^+), the red one is for the bi-atomic oxygen ion (O_2^+), the green one is related to the nitrosonium ion (NO^+) and the cyan dashed line shows the electron density. As mentioned in the previous chapter and as exhibited by Figure 6, the electron density is assumed to be the sum of the other three ions species considered by ANIMo. The graph also shows that the O^+ profile represents the main bulk of the contribution to the ion density and hence the electron density. The profile appears realistic with a peak height around 275 km and a peak density of 10^{12} m^{-3}.

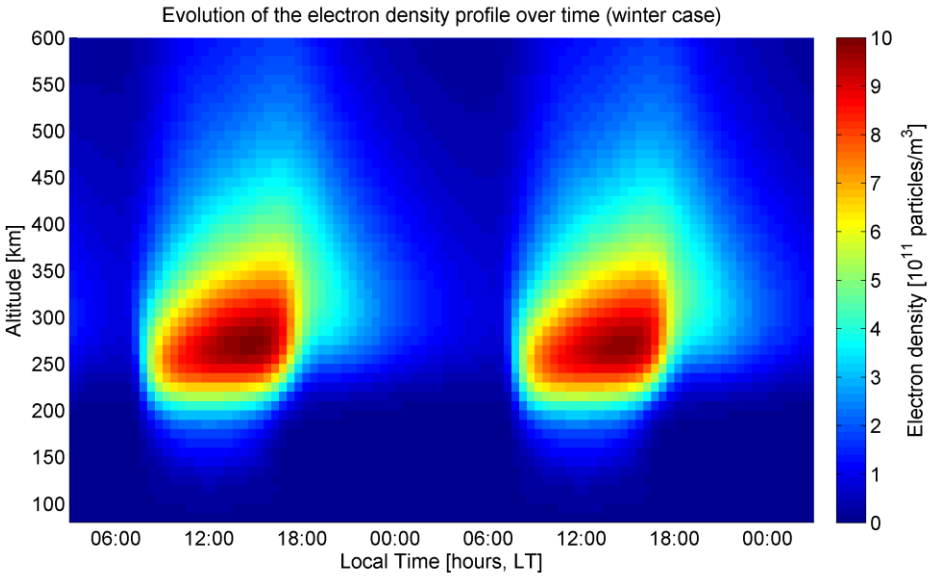

Figure 7. The plot shows the evolution of the electron density profile produced by ANIMo above Millstone Hill Haystack Observatory (Lat. 42.6°, Lon. 288.5°) from 80 to 300 km on the 29th-30th December 2011 (A_p 9-7; *F10.7* 142.3-136.4).

The output of this simulation (Figure 7) is a series of vertical electron density profiles that shows the evolution of the modelled ionosphere. Figure 7 illustrates that ANIMo is capable of reproducing a reasonable day-night plasma variation. In addition to this, it exhibits a rise in the profile peak height at dusk and it possesses a good 'night-time maintenance'. The solar-geomagnetic condition for the winter case, detailed in Table 5, can be seen to be greater on the first day, this is reflected in the model results where the electron density is slightly higher in day 1 with respect to day 2.

Three-dimensional reconstructions in time can be built by combining the evolution of vertical profiles selected from adjacent locations normally defined by means of a grid. By integrating along each profile, it is possible to generate TEC maps of the selected lattice. Several examples of TEC maps produced by ANIMo can be found in Chapter 8, in particular Figure 25 shows the model realization, which for consistency refers to the 29th of December 2011 at 1900 UT, used as background information into the DA scheme. Figure 35 show the evolution in TEC modelled by ANIMo in prediction, hence by not using any contribution by GPS observations.

6.2 ANIMo validation

The validity of the model was tested against different instruments and other ionospheric models. In this document five validation tests are presented, one for each season of the year in geomagnetic quiet times and one for the winter season in unsettled conditions (Subsection 6.2.1). The chosen location is once again Millstone Hill (Lat. 42.6°, Lon. 288.5°); this allowed comparison of the model with measurements from the local ISR and ionosonde. Furthermore, the location was selected in

the past for the inter-comparison of physical models by the Ionospheric-Thermospheric community [*Anderson et al.*, 1998].

A geomagnetic unperturbed period with medium-low solar intensity was chosen for the first four experiments. Regarding the settings of ANIMo input parameters, although the vertical velocity adjustment is kept the same for each case of study, the top boundary condition may be different in order to achieve a greater accuracy.

Table 5 reports in details the aforementioned facts about the selected case studies.

Case study	**Validation test parameters (Input parameters)**			
	Dates	A_p	***F10.7***	**Top boundary condition**
Winter	29-30/12/2011	9 - 7	142.3 - 136.4	default
Spring	09-11/03/2010	2 - 9 - 10	76.8 - 79.3 - 83.1	x2 default
Summer	23-25/06/2011	18 - 11 - 6	99.5 - 99.4 - 96.7	x3 default
Autumn	07-08/09/2010	10 - 11	77.3 - 75.6	x3 default

Table 5. Details about the presented case studies for the validation test. They correspond, together with the selected location (geographic latitude and longitude), to the used input parameters. ANIMo is able to retrieve A_p and *F10.7* parameters automatically.

The following graphs (Figure 8, Figure 9, Figure 10, and Figure 11) display the outcomes of the validation tests. They show the comparison of the evolution in terms of NmF2 and hmF2 between Millstone Hill ionosonde (black line) and ISR measurements (blue), IRI-2012 (red) and ANIMo (green) simulations.

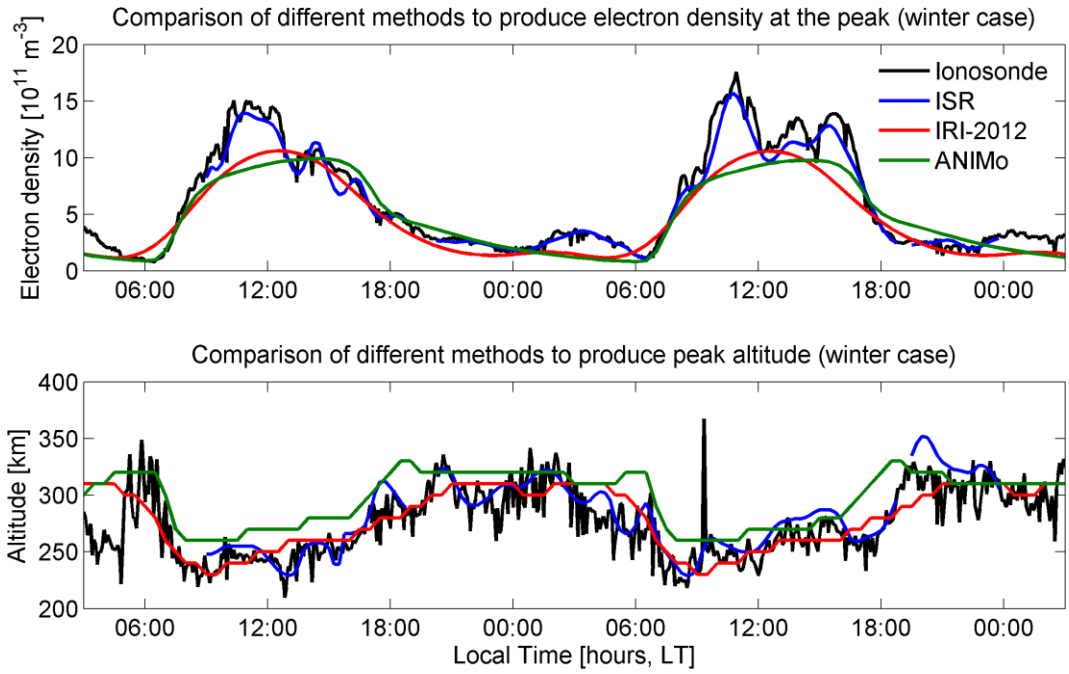

Figure 8. Validity test (winter case). The graphs show, respectively, the comparisons of electron densities at the peak and peak heights measured by Millstone Hill ionosondes and ISR, modelled by IRI 2012 and produced by ANIMo.

Figure 8 shows the comparison of different methods for the winter case. The ionosonde and ISR measurements have a very good agreement for both NmF2. This is due to the fact that the ionosondes was used to calibrate the ISR density. The similarity is confirmed in Table 6, which collects mean, absolute mean and Root Mean Square (RMS) of the difference between the validation criteria (NmF2 and hmF2) of each method against the ionosondes ones. Table 6 displays also similar comparisons by using the ISR hmF2, which it is believed to be more accurate than the ionosonde for this particular criterion.

Winter case				
NmF2 [10^{11} m^{-3}]		**ISR**	**IRI-2012**	**ANIMo**
	Mean of the difference	-0.437	-1.398	-1.143
	Absolute mean of the difference	0.677	1.624	1.649
	RMS of the difference	0.960	2.259	2.357
hmF2 [km]		**ISR**	**IRI-2012**	**ANIMo**
	Mean of the difference	8.552	4.456	22.395
	Absolute mean of the difference	13.471	14.157	24.267
	RMS of the difference	17.683	18.443	28.979
ISR hmF2 [km]			**IRI-2012**	**ANIMo**
	Mean of the difference		0.183	17.494
	Absolute mean of the difference		17.172	21.645
	RMS of the difference		22.851	28.912

Table 6. Validation statistics (winter case). The table reports the mean, absolute mean and RMS of the difference between each method and the ionosonde, for NmF2 and hmF2, and the ISR, for ISR hmF2.

IRI-2012 and ANIMo perform equally well in reproducing the electron density at the peak for this case study. This is very clear in the first graph of Figure 8 (red and green lines) and it is confirmed by the validation statistics describe in Table 6. To be precise, both ANIMo IRI-2012 are slightly underestimating the NmF2 measured by the ionosonde and they struggle in sensing the night-time enhancement visible in the first graph of Figure 8 between 0100 and 0700 LT (0600 and 1200 UT) of day 2 (30th of December 2011). A part from this, ANIMo seems to be able to follow the daily rising and descending of the electron density. Regarding the peak altitude, IRI-2012 is performing better than ANIMo, which tends to overestimate it especially during daytime (second graph in Figure 8). Similarly for the winter case, Figure 9 and Table 7 report results of the validation test for the spring case study.

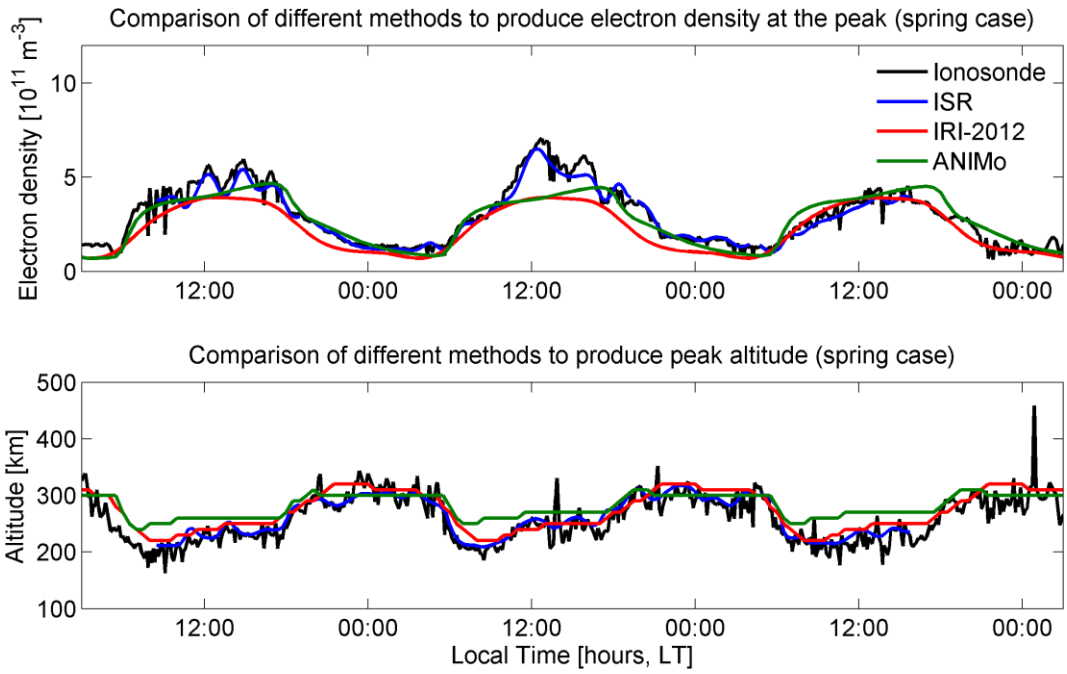

Figure 9. Validity test (spring case). The graphs show, respectively, the comparisons of electron densities at the peak and peak heights measured by Millstone Hill ionosondes and ISR, modelled by IRI 2012 and produced by ANIMo.

As observed in the winter case, ISR NmF2 measurements follow the ionosonde ones closely while ANIMo and IRI-2012 are not capable to perceive the change over the three days. Having said this, statistics from Table 7 tell that ANIMo is performing better than IRI-2012 in the simulation of NmF2. The first graph of Figure 9 shows that even if the models' results are lower than the measurements, ANIMo fits better on the diurnal increase a part for day 3 (11th of March 2010).

Spring case				
NmF2 [10^{11} m^{-3}]		**ISR**	**IRI-2012**	**ANIMo**
	Mean of the difference	-0.175	-0.684	-0.232
	Absolute mean of the difference	0.278	0.741	0.594
	RMS of the difference	0.373	0.961	0.799
hmF2 [km]		**ISR**	**IRI-2012**	**ANIMo**
	Mean of the difference	5.096	13.630	22.044
	Absolute mean of the difference	12.966	19.237	27.278
	RMS of the difference	16.981	23.575	32.241
ISR hmF2 [km]			**IRI-2012**	**ANIMo**
	Mean of the difference		20.676	29.056
	Absolute mean of the difference		23.451	30.394
	RMS of the difference		34.898	38.379

Table 7. Validation statistics (spring case). The table reports the mean, absolute mean and RMS of the difference between each method and the ionosonde, for NmF2 and hmF2, and the ISR, for ISR hmF2.

Peak altitudes values (second graph in Figure 9) are, once more, problematic for ANIMo, which is still overestimating them during daytime. However, ANIMo follows the diurnal trend produced by the other methods. The following comparison refers to the summer case.

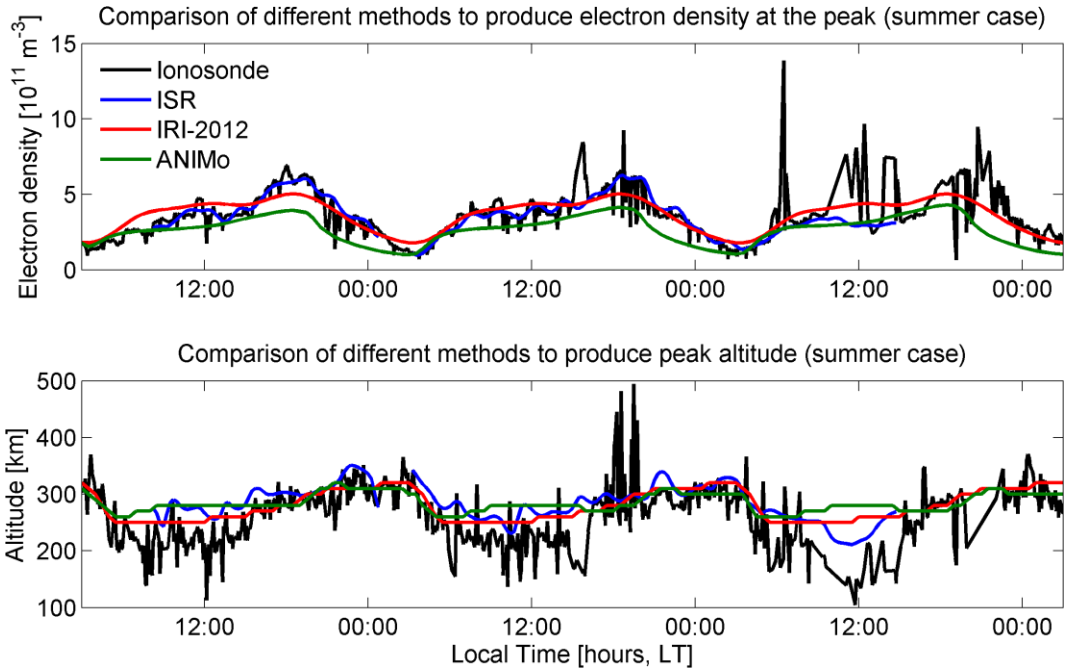

Figure 10. Validity test (summer case). The graphs show, respectively, the comparisons of electron densities at the peak and peak heights measured by Millstone Hill ionosondes and ISR, modelled by IRI 2012 and produced by ANIMo.

Figure 10 shows that the usual good agreement between the ionosonde and the ISR data is sometimes not respected. In particular, ionosonde values are very noisy and discontinuous in the day-time period of day 2 and 3 (24th and 25th of June 2011). This is reflected on the comparison statistics reported in Table 8, where IRI-2012 is shown to perform better than the ISR for both NmF2 and hmF2 values.

Summer case				
NmF2 [10^{11} m^{-3}]		**ISR**	**IRI-2012**	**ANIMo**
	Mean of the difference	-0.260	-0.111	-1.105
	Absolute mean of the difference	0.703	0.824	1.220
	RMS of the difference	1.448	1.403	1.831
hmF2 [km]		**ISR**	**IRI-2012**	**ANIMo**
	Mean of the difference	36.577	24.148	29.114
	Absolute mean of the difference	45.308	36.242	43.493
	RMS of the difference	55.803	48.297	57.714
ISR hmF2 [km]			**IRI-2012**	**ANIMo**
	Mean of the difference		-1.404	4.056
	Absolute mean of the difference		19.588	19.078
	RMS of the difference		24.096	24.386

Table 8. Validation statistics (summer case). The table reports the mean, absolute mean and RMS of the difference between each method and the ionosonde, for NmF2 and hmF2, and the ISR, for ISR hmF2.

By looking at the NmF2 comparison graph of Figure 10, ANIMo underestimates the electron density at the peak during day-time and tends to anticipate the dusk decrease. In contrast, the other methods exhibit more pronounced night-time maintenance. A certain incongruity between methods is very evident by comparing hmF2 values (see hmF2 and ISR hmF2 comparisons in Table 8). As the ionosonde data are less reliable for this case study, it is sensible to compare IRI-2012 and ANIMo results against the ISR measurements. IRI-2012 and ANIMo follow well the diurnal trend but they look more 'static' with respect to the ISR data. ANIMo is also overestimating the day-time peak here, which seems to grow slightly, instead to lie at lower altitude, during day 2 and 3. The last case considered in this section is related to autumn (Figure 11 and Table 9). With regards to NmF2, ionosonde and ISR data are congruent. IRI-2012 and ANIMo behave similarly well in both days (apart a small underestimation at day time) and follow nicely the diurnal trend. Once again, ANIMo tends to overestimate the peak altitude during day-time.

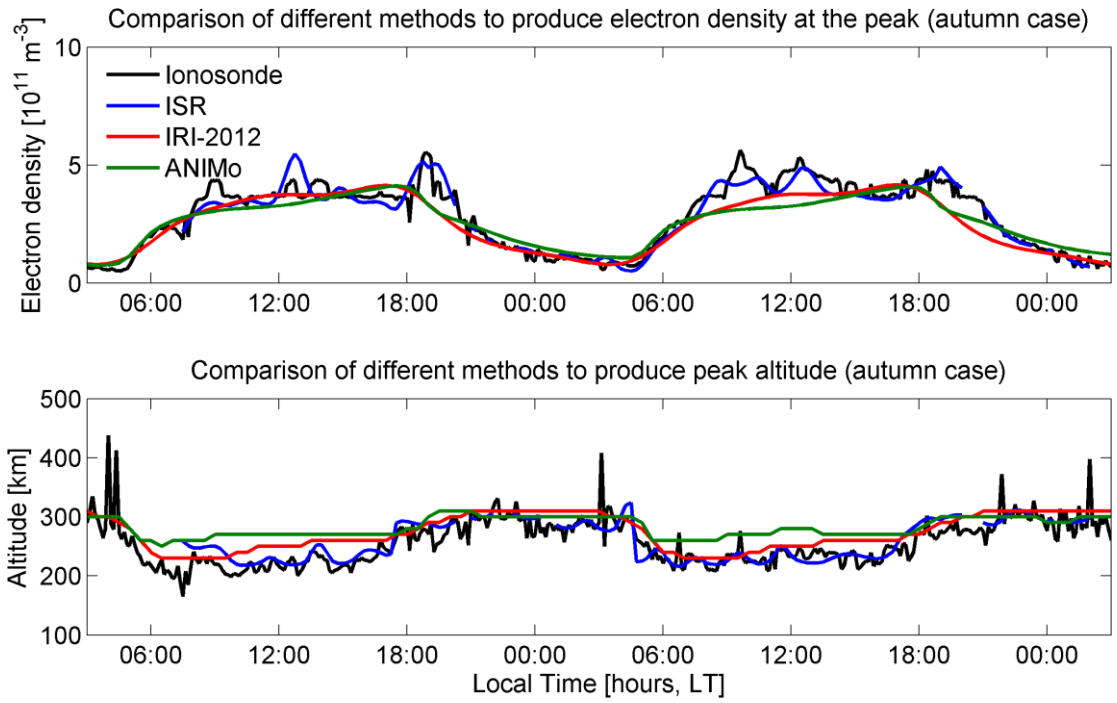

Figure 11. Validity test (autumn case). The graphs show, respectively, the comparisons of electron densities at the peak and peak heights measured by Millstone Hill ionosondes and ISR, modelled by IRI 2012 and produced by ANIMo.

The peak altitude measured by the ISR is very similar to the ionosonde one. The second graph in Figure 11 and Table 9 show that IRI-2012 and ANIMo are capable of simulating, with similar accuracy, the hmF2 of the ionosonde and ISR.

Autumn case				
NmF2 [10^{11} m^{-3}]		**ISR**	**IRI-2012**	**ANIMo**
	Mean of the difference	-0.070	-0.226	-0.188
	Absolute mean of the difference	0.327	0.453	0.526
	RMS of the difference	0.440	0.632	0.703
hmF2 [km]		**ISR**	**IRI-2012**	**ANIMo**
	Mean of the difference	6.573	17.713	26.373
	Absolute mean of the difference	14.769	24.133	33.106
	RMS of the difference	20.853	30.830	40.098
ISR hmF2 [km]			**IRI-2012**	**ANIMo**
	Mean of the difference		11.809	21.921
	Absolute mean of the difference		19.180	25.022
	RMS of the difference		22.384	30.478

Table 9. Validation statistics (autumn case). The table reports the mean, absolute mean and RMS of the difference between each method and the ionosonde, for NmF2 and hmF2, and the ISR, for ISR hmF2.

6.2.1 'Unsettled' winter case study

The validation includes a further case study for the Millstone Hill location. It refers to more perturbed ionospheric conditions with respect to the previous case studies. Table 10 displays the details of this 'unsettled' winter case study.

Case study	Validation test parameters (Input parameters)			
	Dates	A_p	***F10.7***	**Top boundary condition**
Unsettled Winter	04-05-06/02/2011	22 – 16 - 12	79.8 – 78.7 – 78.0	default

Table 10. Details about the unsettled winter case study for the validation test. They correspond, together with the selected location (geographic latitude and longitude), to the used input parameters. ANIMo is able to retrieve A_p and *F10.7* parameters automatically.

The daily A_p index was medium-high level during day 1; this indicates a fairly active geomagnetic situation. Although the solar activity was not particularly high, the 3-hourly K_p indices reached and went over 5 during the last hours of day 1. The effects of the perturbed geomagnetic condition is visible in NmF2 and hmF2 measured by ionosonde and ISR and displayed in Figure 12. The latter shows the comparison of the validation test criteria used so far between different methods for the 'unsettled' winter case study.

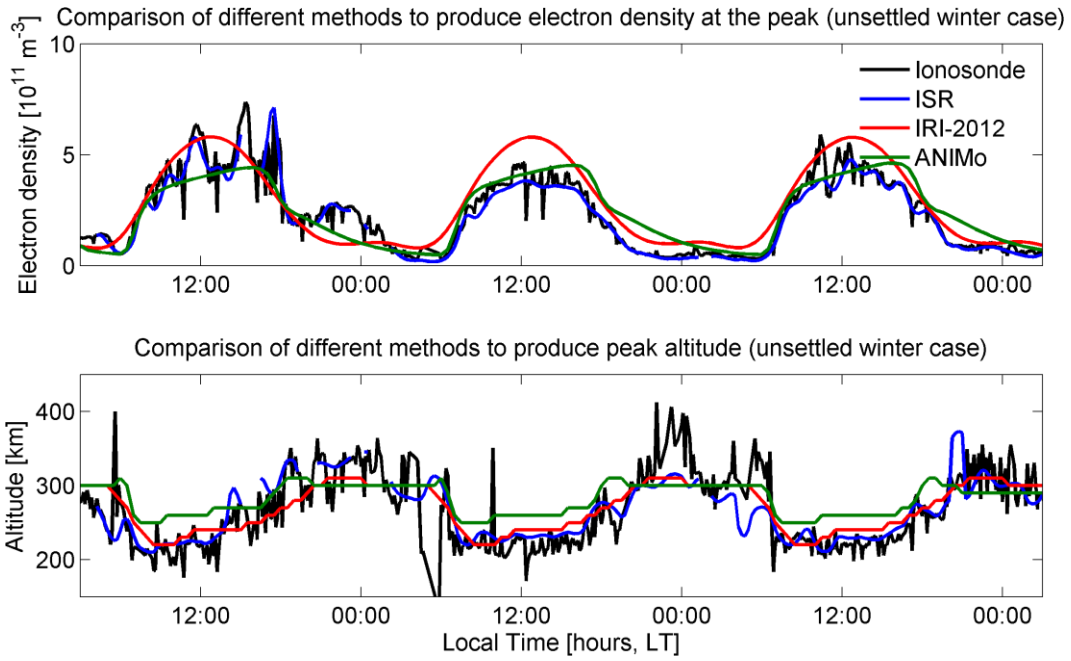

Figure 12. Validity test (unsettled winter case). The graphs show, respectively, the comparisons of electron densities at the peak and peak heights measured by Millstone Hill ionosondes and ISR, modelled by IRI 2012 and produced by ANIMo.

The first graph in Figure 12 shows clearly that the measured NmF2 values for day 1 are very disturbed. It is also present a latent structure at around 2200 LT (0300 UT) that maintains and high NmF2 after dusk of day1. Here ANIMo and IRI do not follow the measured irregularities. On the contrary, the NmF2 evolution in days 2 and 3 looks smoother in agreement with lower

geomagnetic conditions. Although ANIMo tends to overestimate after dusk, it seems more comfortable in reproducing the NmF2 trend during these days. According to the statistics reported in Table 11, ANIMo performs well over the considered period. Regarding the peak altitude ANIMo seems, once again, to be less flexible with respect to the other methods.

'Unsettled' winter case				
NmF2 [10^{11} m^{-3}]		**ISR**	**IRI-2012**	**ANIMo**
	Mean of the difference	-0.171	0.378	0.041
	Absolute mean of the difference	0.310	0.783	0.640
	RMS of the difference	0.496	0.977	0.851
hmF2 [km]		**ISR**	**IRI-2012**	**ANIMo**
	Mean of the difference	-2.790	3.310	15.586
	Absolute mean of the difference	23.180	24.254	33.468
	RMS of the difference	36.010	32.631	40.605
ISR hmF2 [km]			**IRI-2012**	**ANIMo**
	Mean of the difference		7.351	19.565
	Absolute mean of the difference		44.664	42.649
	RMS of the difference		51.113	48.972

Table 11. Validation statistics (unsettled winter case). The table reports the mean, absolute mean and RMS of the difference between each method and the ionosonde, for NmF2 and hmF2, and the ISR, for ISR hmF2.

6.3 Temperature sensitivity

The aim of this sensitivity test was to better understand the changes in electron density that could result from changes to the input temperature. Temperature is a challenging parameter to measure in the ionosphere and is often derived from ISR data. In this model, the temperature is set using the standard run of IRI-2012. If this was far from reality and the model was very sensitive to temperature changes, then it could result in an inaccurate electron density.

In particular, this paragraph reports a selection from a series of tests conducted by tuning the temperature input parameter. The chosen case study is that of winter, as already presented in the validation test. Figure 13 shows the comparison between outcomes obtained by using different temperature input values. As aforementioned, ANIMo normally uses temperature values produced by IRI-2012, the relative outcome of which is reported in the graph with a solid green line for consistency with previous figures. The model was also fed with temperature measurements from the Millstone Hill ISR (blue solid) and artificial profiles defined by keeping the temperature constant in altitude and time at 1000 K (gold dashed), 2000 K (orange dashed) and 3000 K (red dashed). In addition to Figure 13, Table 12 reports mean, absolute mean and RMS of the difference between NmF2 and hmF2 of each temperature input set-up (ISR temperature, constant 1000 K, 2000 K and 3000 K) against the canonical ANIMo. The test demonstrates the importance of ion

and electron temperatures as input in modelling the electron density of the ionosphere. Furthermore, it shows that ANIMo is a robust model in terms of temperature modification, where robustness is defined as the ability of coping with large changes of external forcing parameters.

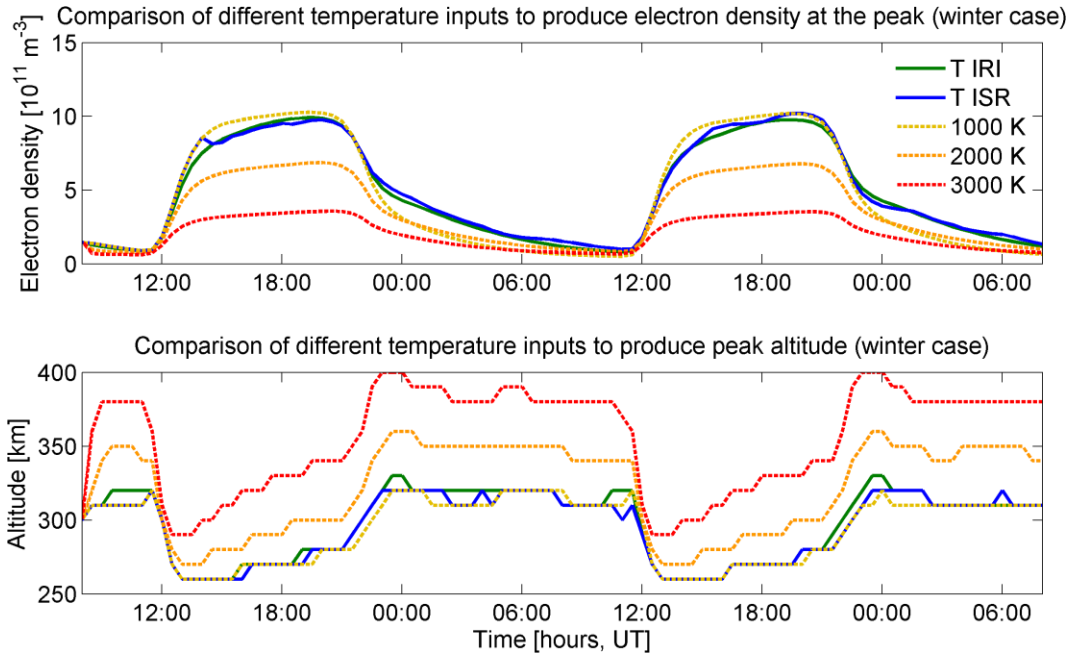

Figure 13. Sensitivity test statistics for temperature (winter case). The graphs show, respectively, the comparisons of electron densities at the peak and peak heights produced by modifying ANIMo temperature input parameters.

Winter case (Temperature sensitivity)					
NmF2 [10^{11} m^{-3}]		**T ISR**	**1000 K**	**2000 K**	**3000 K**
	Mean of the difference	-0.133	0.209	1.360	2.844
	Absolute mean of the difference	0.225	0.665	1.362	2.845
	RMS	0.285	0.774	1.721	3.637
hmF2 [km]		**T ISR**	**1000 K**	**2000 K**	**3000 K**
	Mean of the difference	2.165	3.196	-24.330	-57.423
	Absolute mean of the difference	2.371	3.608	24.330	57.423
	RMS	5.077	6.176	25.966	59.439

Table 12. Sensitivity test statistics for temperature (winter case). The table reports the mean, absolute mean and RMS of the difference between each temperature input set-up and the canonical ANIMo (temperature provided by IRI-2012).

IRI-2012 and ISR temperature-driven outcomes are very similar (green and blue solid lines in Figure 13). The statistics shown in Table 12 confirm this similarity; T ISR values are in fact very low for both comparison criteria. Regarding the remaining simulations, increasing the selected input value translates to a gradual increase of the alteration of the model results (Table 12). As expected, the higher the temperature, the smaller the electron density and the bigger the peak altitude. This is due to the fact that temperature affects the recombination rates and diffusion velocities in ANIMo. In particular, if the recombination rate increases there will not only be a

general decrease in electron and ion densities but also a lift of the peak altitude that is not replenished enough by the photoionization. In addition to this, the collision frequency is greater in a hotter environment. This fact, plus the diminished charged particle density slows down the diffusion that tends to move ions and electrons to lower positions of the profile.

6.4 Top-side flux sensitivity

The top-side boundary conditions is one of the ANIMo's inputs that needs to be defined by the user. The sensitivity test reported in this section aims to show how the model output changes by tuning this parameter.

Five different values of downward fluxes were selected and used as top-side boundary conditions in reproducing the Millstone Hill winter case. The chosen fluxes were kept constant during each model run. Figure 14 shows a comparison between realizations of ANIMo obtained by using the mentioned flux parameters. In particular, the solid green line refers to the default value used for the winter case. The dashed black line refers to 0 $m^{-2}s^{-1}$, the light blue to the factor of two (x2) of the default value, the medium blue line to x3 of the default and the navy blue one to x4 of the default. Similarly to the temperature sensitivity test, the canonical winter ANIMo reconstruction is used as reference to generate mean, absolute mean and RMS of the difference between NmF2 and hmF2 with each other flux inputs. The statistics results are reported in Table 13.

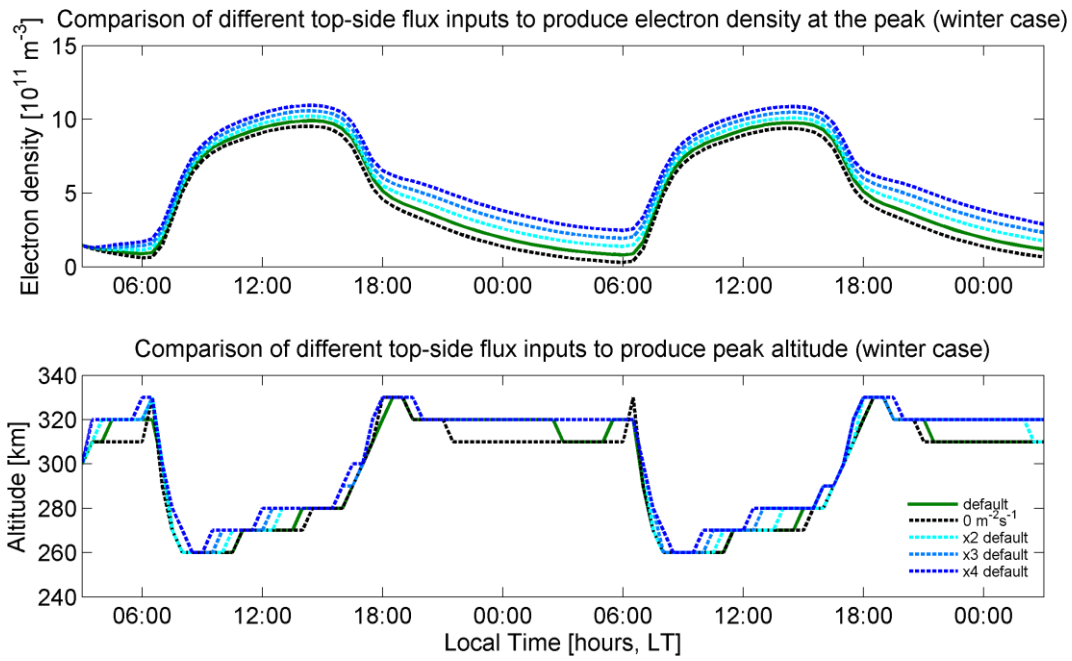

Figure 14. Sensitivity test for the top-side flux (winter case). The graphs show, respectively, the comparisons of electron densities at the peak and peak heights produced by modifying ANIMo top boundary conditions.

Winter case (Top-side flux sensitivity)					
NmF2 [10^{11} m^{-3}]		**Flux 0**	**Flux x2**	**Flux x3**	**Flux x4**
	Mean of the difference	0.457	-0.428	-0.883	-1.344
	Absolute mean of the difference	0.457	0.429	0.884	1.349
	RMS	0.477	0.464	0.940	1.424
hmF2 [km]		**Flux 0**	**Flux x2**	**Flux x3**	**Flux x4**
	Mean of the difference	1.856	-2.990	-4.330	-5.155
	Absolute mean of the difference	2.474	2.990	4.330	5.155
	RMS	4.974	5.468	6.580	7.180

Table 13. Sensitivity test statistics for the top-side flux (winter case). The table reports the mean, absolute mean and RMS of the difference between each input flux value (expressed in factors of the default value) and the ANIMo winter default top boundary condition.

Figure 14 and Table 13 show clearly that introducing a higher downward flux increases the NmF2 values particularly during night time. This is due to the diffusion velocity adjustment that is enabled only during day time (Section 5.2.2) to guarantee the night-time maintenance. A sensitivity test for the vertical velocity adjustment is described in Section 6.5. The variation of the flux parameter does not seem to considerably affect the peak altitude; however the rising of the hmF2 during daytime tends to be anticipated for bigger values of top-boundary flux.

6.5 Vertical velocity adjustment sensitivity

The vertical diffusion velocity generated by ANIMo can be adjusted in order to produce more accurate reconstructions of the ionosphere. As explained in Subsection 5.2.2, the vertical velocity adjustment is enabled only during day-time and decreases along the vertical profile by following a Gaussian function. Similarly for the top-boundary flux, a sensitivity test was performed to demonstrate the effects of playing with this input parameter. The test results are displayed in Figure 15 in terms of NmF2 and hmF2. Once again, the green solid line refers to the default settings of ANIMo for the winter case study over Millstone Hill. The model was run for the same case by setting the vertical velocity to 0 m s^{-1} (dashed black line) as well as to half of the default value (dashed magenta line), and to 3/2 of the default (dashed purple line).

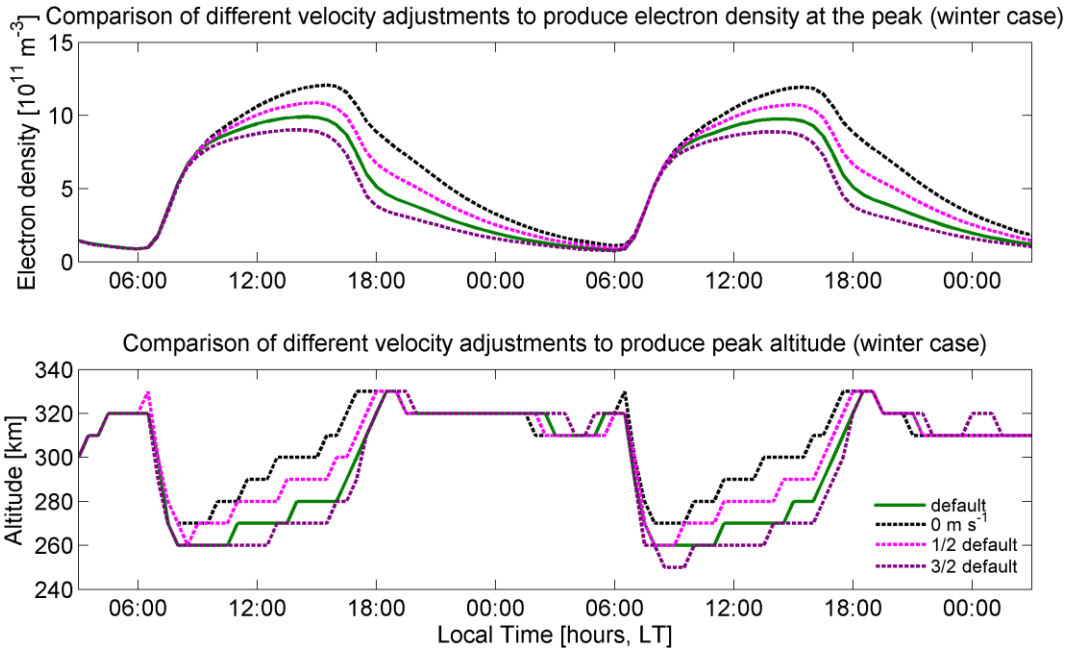

Figure 15. Sensitivity test for the vertical velocity adjustment (winter case). The graphs show, respectively, the comparisons of electron densities at the peak and peak heights produced by modifying ANIMo vertical velocity adjustment.

Table 14 reports the test statistics that were calculated by taking the ANIMo run with the default velocity adjustment as reference.

Winter case (Vertical velocity adjustment sensitivity)				
NmF2 [10^{11} m^{-3}]		**Adj. 0**	**Adj. x1/2**	**Adj. x3/2**
	Mean of the difference	-1.429	-0.636	-0.549
	Absolute mean of the difference	1.429	0.645	0.552
	RMS	1.844	0.808	0.697
hmF2 [km]		**Adj. 0**	**Adj. x1/2**	**Adj. x3/2**
	Mean of the difference	-8.660	-4.336	1.856
	Absolute mean of the difference	9.484	4.742	3.505
	RMS	14.142	7.322	5.920

Table 14. Sensitivity test statistics for the vertical velocity adjustment (winter case). The table reports the mean, absolute mean and RMS of the difference between each velocity adjustment value (expressed in factors of the default adjustment) and the ANIMo default one.

The effect of the introduction of the velocity adjustment is very clear from the results of this test. Although the rising of NmF2 at dawn is unaffected, a higher adjustment value curbs the evolution during day and especially after dusk. In particular, the greater is the velocity and the faster is the decrease of electron density peak during evening and night-time. On the contrary, the peak altitude is not altered in the night-time and the greater is the velocity the later the hmF2 rising happens. This behaviour is simply due to the fact that the daily vertical adjustment contrasts the upward plasma diffusion of the higher portions of the ionosphere. The result is that ions are pushed in the lower regions of the profile where the recombination is stronger. This tends not only to decrease

the plasma density of the profile but also to keep the peak at a lower altitude for a longer time. The peak will eventually rise again when the day-time photoionization generates enough plasma to contrast both adjustment and recombination.

Summary

The testing stages of ANIMo were fundamental to comprehend whether (and when) ANIMo was ready and suitable to be used in support of ionospheric tomography. As reported in the previous chapter in Section 5.1, the model needed to meet specific requirements to fulfil the general project objectives. Accuracy is surely one of the most relevant; the validation demonstrated that ANIMo is capable of reproducing different features of the ionosphere in a reasonable manner, considering the physics that had been taken into account. This was confirmed by previous comparisons with the Utah State University Time Dependent Ionospheric Model (USU TDIM) [*Sojka et al.*, 2013].

Validation results presented in this chapter show that the model simulation follows reasonably day-night evolution of the electron density at the peak and peak altitude. Although it is not capable to reproduce disturbed ionospheric features, the model performs reasonably also in medium unsettled conditions (Subsection 6.2.1). The selection of different case studies from different periods of the year indicates that ANIMo is also sensitive to seasonal variations. Averages of the NmF2 and hmF2 provided by the different methods (Millstone Hill ionosonde and ISR measurements, IRI-2012 and ANIMo simulations) were calculated for each unperturbed case study, hence for each season, and compared in Figure 16. IRI-2012, because of its empirical nature (see Section 4.1.1), is here the best reference of comparison.

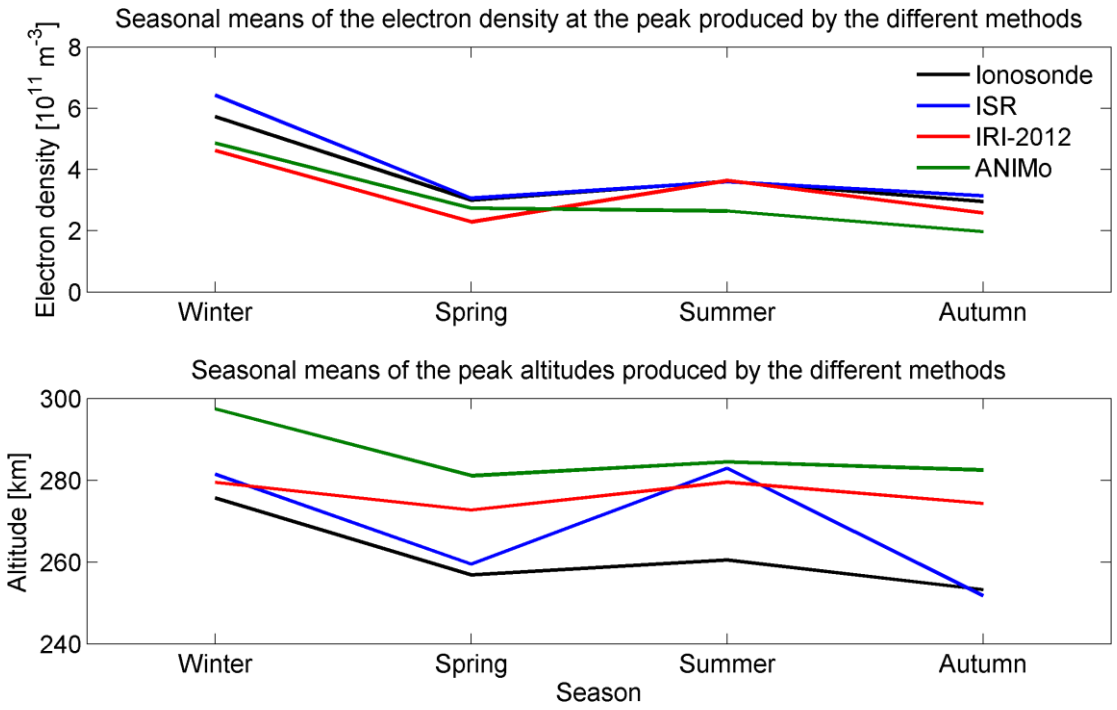

Figure 16. Seasonal performances comparison. The graphs show, respectively, the comparisons of the means of electron densities at the peak and peak heights measured by Millstone Hill ionosondes and ISR, modelled by IRI 2012 and produced by ANIMo for each case study (season).

The first graph in Figure 16 shows that, although there is a sensible underestimation in summer, ANIMo (green solid line) is capable of following the annual trend of the mean of NmF2 sensed by the other methods. The second graph displays the same kind of comparison but related to the average hmF2. ANIMo tends to overestimate this criterion in each season; however, its seasonal behaviour is remarkable. An additional interesting result comes from the testing of ANIMo. The fact that this model is able to sense substantial differences in electron density (and their altitudes) over the four case studies can give some indications about the physical processes involved in ionospheric seasonal effects. This is possible by exploiting the non-complexity feature of ANIMo, for which small amount of input parameters need to be defined in order to have a fairly accurate simulation. The elements in ANIMo that change in each case study are: input parameter as top-side boundary condition, the geometry of the solar incident ray-path, the neutral composition provided by MSIS and the ion and electron temperature estimated by IRI-2012. Although more investigation is needed, the temperature seems to have a key role, with respect to the mentioned elements, in the seasonal change and, in particular, in the winter anomaly. The magnitude of its influence can be appreciated in Figure 13 and Table 12 in Section 6.3. Similarly, the effects of tuning the top-side boundary condition and the vertical velocity adjustment, which has not been changed in all the presented experiments, are shown in Figure 14 and Table 13 in Section 6.4, and Figure 15 and Table 14 in Section 6.5 respectively. Table 15 summarises how ANIMo's output get affected by increasing the aforementioned input parameters.

Effects on ANIMo output		
	NmF2	**hmF2**
Higher temperature	Lower	Higher
Higher top-side boundary flux	Higher (esp. night-time)	Earlier day-time rising
Higher vertical velocity adjustment	Lower	Later day-time rising

Table 15. Effects on ANIMo outputs by increasing the value of temperature, top-side boundary flux and vertical velocity adjustment.

This is a useful guideline in order to reach a good accuracy. ANIMo results could be therefore improved by including a better representation of these parameters.

By pushing the input parameters to the extreme, especially the temperature one, it has been possible to demonstrate that ANIMo is capable to expectably deal with them. Being robust and reliable are very important qualities considering that the model is expected to be used to simulate specifically unsettled conditions and study their evolution. Although not reported, a considerable amount of tests were performed in order to increase the efficiency of ANIMo. In particular, they were successfully applied in reducing the time of execution of programming codes. At this phase of its development, ANIMo is hence exhibiting the characteristics required for supporting ionospheric tomography imaging. Tests were also fundamental to comprehend the limitations of the model and the path that can be taken for its future expansions.

Chapter 7 A New Ionospheric Data Assimilation System (ANIDAS)

Abstract

This chapter develops and applies the theory to use the model ANIMo in Data Assimilation. A New Ionospheric Data Assimilation System (ANIDAS) is a data ingestion approach for the upper atmosphere. Based on a variational DA scheme, it combines background information provided by ANIMo and GPS observations to produce a more accurate representation of the ionosphere.

The inputs are time and location (defined by a three-dimensional grid) of the desired reconstruction. GPS observations are RINEX files gathered from various on-line repositories and pre-processed in order to have STEC information and to construct a projection matrix. The observations are collected over a period of time divided in multiple time frames. The whole time window is centred on the time of the reconstruction defined as an input. A model realization from ANIMo is generated for this time and used as background; the latter is assumed to be constant during the entire observation time window. ANIMo is also used to create vertical basis functions by means of two different techniques. The first one adopts the background information to constrain directly the shape of each vertical profile in the grid. The second one creates the constraints starting from an ensemble of possible realizations of each profile by tuning ANIMo forcing parameters. The observation covariance matrix is assumed to be diagonal, while the background covariance matrix is generated by an internal time-dependent model based on historical data series of electron density correlation distances. Since ANIDAS is based on a variational data assimilation scheme, it can be used iteratively and eventually perform ionospheric forecasting. This is done by initializing the model ANIMo at the current time step with the result of the scheme from the previous time step. ANIMo, strengthened in accuracy, can be exploited to predict the evolution of the current ionosphere.

Introduction

The ultimate aim of ANIMo is to provide information about the ionosphere that is not currently readily available from ionospheric instruments. Thus, it is intended to be useful to augment observations to provide a 3D time dependent representation of the ionosphere, through tomographic imaging or DA. Further, the nature of the model enables the forward projection in time to forecast the ionosphere ahead, or even to run as a stand-alone model without observations.

This chapter describes the design, the function and the main elements of ANIDAS. In particular, section 7.1 is a theoretical explanation of how ANIDAS was developed starting from conventional data ingestion approaches algorithms and MIDAS. This section includes a paragraph (7.1.1) that focuses on the calibration issues. Section 7.2 briefly shows the numerical implementation of ANIDAS and it describes two important elements of the operation. Subsection 7.2.1 explains the background covariance matrix, its meaning, how is generated and how it is applied. Similarly, subsection 7.2.2 illustrates the role of physics-based vertical basis functions within the scheme. ANIDAS can be used in an iterative manner for now-casting and forecasting; this will be expanded in Section 7.3. Subsection 7.3.1 is a brief digression about using different methods for generating vertical basis functions within ANIDAS in now-casting.

7.1 The concept of ANIDAS

ANIDAS can be defined as Four-Dimensional Variational analysis (4DVar). Largely used in NWF, these algorithms are based on the minimization of a cost function, normally referred as $\boldsymbol{J(x)}$:

$$J(\boldsymbol{x}) = (\boldsymbol{y} - \mathbf{H}[\boldsymbol{x}])^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mathbf{H}[\boldsymbol{x}])^{\mathrm{T}} + (\boldsymbol{x} - \boldsymbol{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{x} - \boldsymbol{x}_b) \tag{7.1}$$

Where $\boldsymbol{x}$ is the state vector, $\boldsymbol{y}$ is the vector of the observations, $\mathbf{H}$ is the observation operator, $\boldsymbol{x}_b$ is the background or *a priori* information, $\boldsymbol{\Sigma}$ and $\mathbf{B}$ are the observation and background covariance matrices respectively.

The cost function $\boldsymbol{J(x)}$ can be rewritten as follows:

$$J(\boldsymbol{x}) = \|\boldsymbol{y} - \mathbf{H}[\boldsymbol{x}]\|^2_{\boldsymbol{\Sigma}^{-1}} + \|\boldsymbol{x} - \boldsymbol{x}_b\|^2_{\mathbf{B}^{-1}} \tag{7.2}$$

The solution of the minimization is called the analysis. This technique can be found in greater detail and well explained, in the work of *Bouttier and Courtier* [2002].

ANIDAS was developed from MIDAS algorithms; essentially, it can be viewed as a modification of the latter. As aforementioned MIDAS is a tomography software package. Its core is the mathematical treatment of an inverse problem aided by means of a Tikhonov regularisation [*Tikhonov et al.*, 1977]:

$$J_M(\boldsymbol{x}) = \|\boldsymbol{z} - \mathbf{H}\mathbf{K}\boldsymbol{x}\|^2 + \lambda\left\|\boldsymbol{\nabla}^{\mathbf{2}}\mathbf{K}\boldsymbol{x}\right\|^2 \tag{7.3}$$

Where $\boldsymbol{J_M}$ is the cost function of MIDAS with respect to the state vector $\boldsymbol{x}$, $\mathbf{K}\boldsymbol{x}$ refers to electron density values, $\boldsymbol{z}$ are the GPS observations, $\mathbf{H}$ is the observation operator, λ is a tuning parameter and $\boldsymbol{\nabla}^{\mathbf{2}}$ indicates the usage of a second order Laplacian matrix as regularisation. The observation operator $\mathbf{H}$, in this case, is a matrix that collects the lengths of the intersections of the GPS observation paths with a defined grid that spatially discretises the selected area. In MIDAS, $\mathbf{K}$ is a matrix that contains vertical basis functions and it is used to aid the inversion for improving the

vertical reconstruction. The state vector $\boldsymbol{x}$ therefore consists of coefficients that weight the respective basis function. Electron density values $\boldsymbol{n}$ are hence defined as:

$$\boldsymbol{n} = \mathbf{K}\boldsymbol{x} \tag{7.4}$$

Thus, MIDAS recreates an ionosphere that is a balance between the observations and a regularising based on a second derivative matrix. This is an adequate approach for regions of reasonable data coverage but it is not good enough for representing the global-scale ionosphere where the problem transitions from dense coverage over the continental land regions to sparse or no observations over the oceans. To improve these limitations, it would be useful to have a coherent mathematical approach that will rely more on data where it is available and more upon the model where it is not.

A further limitation is given by the poor capability of resolving vertical profiles due to the geometry of the satellite rays. Basis functions generated by ANIMo can be used to form a useful mathematical frame to represent the vertical ionospheric profile and, where more information becomes available, they can easily allow for changes to the vertical profile shape. Section 7.2.2 will thoroughly explain how basis functions are generated, and their role in the inversion process.

Referring back to Equation (7.4) it is noted that the regularisation $\left\|\boldsymbol{\nabla^2 \mathbf{K} x}\right\|^2$ is based on the *a priori* assumption that very sharp electron density gradients are not present in the ionosphere. The parameter $\boldsymbol{\lambda}$ decides the importance of the contribution of the regularisation. If $\boldsymbol{\lambda}$ is very small, the whole operation relies mostly on observations and hence the final result may be affected by artefacts, where there are no data, for the presence of singularities in the inversion operator. By increasing $\boldsymbol{\lambda}$ the inversion becomes more stable, thus less noisy but possibly in less close agreement to the observations. There is a risk of using a very big $\boldsymbol{\lambda}$ which will filter out real small scale details and ignore strong gradients that the observations were able to resolve. Effects of the tuning of $\boldsymbol{\lambda}$ are shown in the Appendix A (Figure 41 collects the pictures that show MIDAS alone for the calibration test). To further complicate the matter the whole problem is also time dependent and this also requires incorporation into the inversion. If it were not so, the inversion would have to use short data segments over say, 15 minute intervals (over which time the ionosphere could perhaps be considered to not change) and the short segments would affect the stability and lack the length of segment needed to resolve real features in the ionosphere.

MIDAS collects the necessary observations $\boldsymbol{z}$ over multiple windows of time, centred on the time of the analysis. A full explanation of MIDAS algorithms is provided by *Mitchell and Spencer* [2003]; [*Spencer and Mitchell*, 2007]. By comparing Equations (7.2) and (7.3), the similarity between variational analysis and the inversion problem stands out.

ANIDAS, by implementing ANIMo as background, can be seen as a development of MIDAS towards a full physically driven 4DVar scheme (Figure 17 shows how the scheme components of

ANIDAS contribute to the realization of the analysis). The presence of a first principles model is not the only difference. ANIDAS solves its cost function J_A with respect to a residual state vector $\delta\boldsymbol{x}$. The pseudo-solution $\delta\hat{\boldsymbol{x}}$, which is not necessarily equal to $\delta\boldsymbol{x}$, is called the analysis increment (or gain) and corresponds to the correction to apply to ANIMo background state $\boldsymbol{n}_b$ in order to match the analysis $\hat{\boldsymbol{x}}$ (that can be translated in electron density by means of Equation (7.4)):

$$\hat{\boldsymbol{n}} = \boldsymbol{n}_b + \mathbf{K}\delta\hat{\boldsymbol{x}} \tag{7.5}$$

Where $\hat{\boldsymbol{n}}$ is the analysis in terms of electron density.

The background state $\boldsymbol{n}_b$ is multiplied by the projection matrix $\mathbf{H}$ (linear operator), obtained from the satellite observations $\boldsymbol{z}$ in order to generate the 'modelled observation' $\boldsymbol{z}_b$:

$$\boldsymbol{z}_b = \mathbf{H}\boldsymbol{n}_b \tag{7.6}$$

Similarly to MIDAS, observations $\boldsymbol{z}$ are normally collected at multiple time steps over a defined window of time; for this reason ANIDAS falls within the definition of sequential 4DVar.

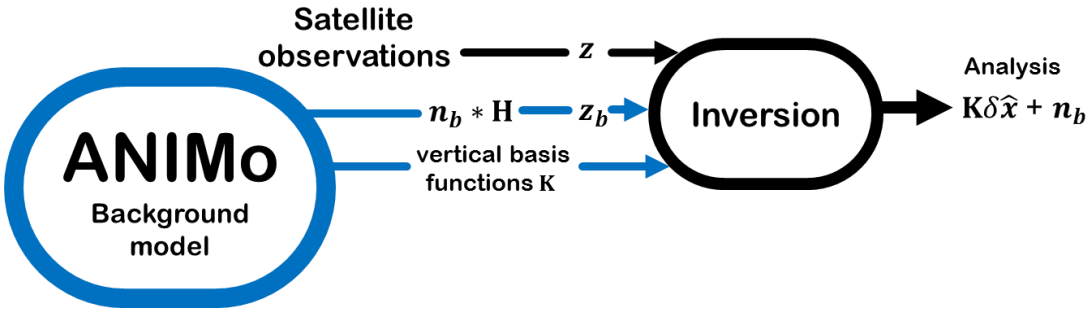

Figure 17. Simplified diagram of the principal components of ANIDAS scheme used in a non-iterative manner.

The observations are then combined to represent the selected time step, which, also for MIDAS, is usually the central one in the time window. The difference between $\boldsymbol{z}_b$ and $\boldsymbol{z}$ gives the observation residual $\delta\boldsymbol{z}$:

$$\delta\boldsymbol{z} = \boldsymbol{z} - \boldsymbol{z}_b \tag{7.7}$$

The cost function J_A, which refers to the inversion block displayed in Figure 17, is therefore written:

$$J_A(\delta\boldsymbol{x}) = \|\delta\boldsymbol{z} - \mathbf{H}\mathbf{K}\delta\boldsymbol{x}\|^2_{\boldsymbol{\Sigma}^{-1}} + \alpha\|\mathbf{K}\delta\boldsymbol{x}\|^2_{\mathbf{B}^{-1}} \tag{7.8}$$

Where $\boldsymbol{\Sigma}$ and $\mathbf{B}$ are the observation and background covariance matrix respectively. It can be seen from the equation that there is a balance to be made between minimising the difference between the observations and the projected state vector $\delta\boldsymbol{x}$, and minimising the difference between the model and the state vector $\mathbf{K}\boldsymbol{x}$. This is not a simple task because the observations and the model have very different types of error. The observations are relative STEC, with negligible error, and

the model contains information on the vertical ionospheric profile and is anticipated to be rather poor in representing STEC. This topic of error estimation will be returned to throughout the chapter.

While the observation covariance matrix $\mathbf{\Sigma}$ is for convenience assumed diagonal, the background covariance matrix $\mathbf{B}$ is generated by a parallel empirical model. The latter was also developed during the doctoral project and it is fully explained in section 7.2.1. The parameter α is similar to λ and controls the background contribution and its mathematical meaning is related to $\mathbf{B}$. This concept is expanded in section 7.2.1.

7.1.1 Feasibility of using MIDAS algorithms for inverting residual values: The calibration dilemma

One of the major concerns in applying the method expanded upon in the previous section is related to how the observation offsets are handled, in other words the calibration. MIDAS uses un-calibrated data and performs a calibration approach which is internal in its inversion algorithms. This consists of time-wise differencing observation values from the same satellite-receiver ray; this means that GPS observations in MIDAS are actually relative values. The problem of using the inversion facilities of MIDAS, for ANIDAS, could come from the inadequacy of the calibration processes applied to residual values. The question is therefore: can ANIDAS assure the calibration of the observation offsets?

It is possible to demonstrate that ANIDAS can be rewritten in a non-residual form. For this demonstration, the residual solution (equivalent to $\delta\hat{\boldsymbol{x}}$ from Equation (7.5)) will be called $\delta\hat{\boldsymbol{x}}_r$ and the non-residual $\hat{\boldsymbol{x}}_{nr}$.

According to Equation (7.8), it can be asserted that:

$$\delta\hat{\boldsymbol{x}}_r = \underset{\delta\boldsymbol{x}_r}{\arg\min} J_{A\text{-}r}(\delta\boldsymbol{x}_r) \tag{7.9}$$

In other words, the state vector for the residual version of ANIDAS $\delta\hat{\boldsymbol{x}}_r$ is given by minimizing the ANIDAS residual cost function $J_{A\text{-}r}(\delta\boldsymbol{x}_r)$ that is equivalent to Equation (7.8). Therefore, similarly for Equation (7.5), it is possible to write:

$$\hat{\boldsymbol{n}}_r = \boldsymbol{n}_b + \mathbf{K}\delta\hat{\boldsymbol{x}}_r \tag{7.10}$$

Where $\hat{\boldsymbol{n}}_r$ is the residual analysis in terms of electron density, $\boldsymbol{n}_b$ is the background state from ANIMo and $\delta\hat{\boldsymbol{x}}_r$ is the residual state vector. The non-residual version of Equation (7.8) can be rewritten by using the relations described in Equation (7.5) and (7.7):

$$J_{A\text{-}nr}(\boldsymbol{x_{nr}}) = \|(\boldsymbol{z} - \boldsymbol{z}_b) - \mathbf{H}(\mathbf{K}\boldsymbol{x_{nr}} - \boldsymbol{n}_b)\|^2_{\mathbf{\Sigma}^{-1}} + \alpha\|\mathbf{K}\boldsymbol{x_{nr}} - \boldsymbol{n}_b\|^2_{\mathbf{B}^{-1}} \tag{7.11}$$

Where $J_{A\text{-}nr}$ is the non-residual cost function with respect to $\boldsymbol{x_{nr}}$. Because of Equation (7.6), it is possible to find:

$$J_{A\text{-}nr}(\boldsymbol{x_{nr}}) = \|\boldsymbol{z} - \mathbf{H}\mathbf{K}\boldsymbol{x_{nr}}\|^2_{\mathbf{\Sigma}^{-1}} + \alpha\|\mathbf{K}\boldsymbol{x_{nr}} - \boldsymbol{n}_b\|^2_{\mathbf{B}^{-1}} \tag{7.12}$$

This coincides with the variational formulation described by Equation (7.2) re-expressed with a different notation. Similarly for Equation (7.9), it is possible to write:

$$\hat{\boldsymbol{x}}_{nr} = \underset{\boldsymbol{x_{nr}}}{\arg\min} J_{A\text{-}nr}(\boldsymbol{x_{nr}}) \tag{7.13}$$

Note that, considering Equation (7.4):

$$\hat{\boldsymbol{n}}_r = \hat{\boldsymbol{n}}_{nr} \tag{7.14}$$

This brief demonstration shows that the non-residual formulation of ANIDAS (Eq. (7.8)) coincides with the canonical one of the variational analysis (Eq. (7.2)). As it was mentioned in the previous section, the latter is in turn analogous to the formulation of MIDAS (Eq. (7.3)). Due to the similarity of the optimisation problems, it can be inferred that if MIDAS can estimate offsets so can ANIDAS.

7.2 The implementation of ANIDAS

Equation (7.8) describes the fundamental formula used in ANIDAS. The latter can be expanded to the following formulation which is more similar to the operation performed in ANIDAS algorithm to find the solution.

$$\delta\hat{\boldsymbol{x}} = (\mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{H}\mathbf{K} + \alpha\mathbf{K}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{K})^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\delta\boldsymbol{z} \tag{7.15}$$

Two very important elements are the background covariance matrix **B** and the vertical basis functions matrix **K**.

7.2.1 Background Covariance Matrix

The brief introduction regarding the theory of the background covariance matrix is inspired to the explanation taken form the work of *Bouttier and Courtier* [2002].

The usage of DA techniques requires the statistical awareness of the uncertainty of the background. Considering $\boldsymbol{x}_t$ as the true value that is needed to be estimated, also known as true state, it is possible to assert:

$$\varepsilon_{\mathrm{b}} = \boldsymbol{x}_t - \boldsymbol{x}_b \tag{7.16}$$

The term ε_{b} is then a misestimating of the background $\boldsymbol{x}_b$ in trying to simulate the real value. By repeating this calculation in the same conditions and adding a source of error due to unknown causes, ε_{b} would change each time. From the analysis of the statistics of ε_{b} generated by a great

number of repetitions, it would be possible to observe a dependence to the physical processes responsible for those errors. The statistical awareness of this error is carried in the DA algorithm by the background covariance matrix **B**. The latter, in a scalar system, corresponds to the variance of ε_b:

$$\mathbf{B} = \overline{(\varepsilon_b - \overline{\varepsilon_b})^2} \tag{7.17}$$

In a multidimensional system, **B** becomes a squared matrix whose order is equal to the number of elements in the state vector. The main diagonal is the vector of all the variances of ε_b of the system while the off-diagonal elements are filled with the covariance values between every combination of two different variables in the same system. The background covariance matrix **B** for a tri-dimensional system can be drawn as follows:

$$\mathbf{B} = \begin{bmatrix} \mathrm{var}(e_1) & \mathrm{cov}(e_1, e_2) & \mathrm{cov}(e_1, e_3) \\ \mathrm{cov}(e_2, e_1) & \mathrm{var}(e_2) & \mathrm{cov}(e_2, e_3) \\ \mathrm{cov}(e_3, e_1) & \mathrm{cov}(e_3, e_2) & \mathrm{var}(e_3) \end{bmatrix} \tag{7.18}$$

Where for each dimension i, the value e_i is the difference between the background error ε_{bi} and its average $\overline{\varepsilon_{bi}}$. Covariance values can be translated into correlations, by considering the following relation (and the generic dimension j):

$$\rho(e_i, e_j) = \frac{\mathrm{cov}(e_i, e_j)}{\sqrt{\mathrm{var}(e_i)\mathrm{var}(e_j)}} \tag{7.19}$$

For real applications, it is not possible to calculate the covariance matrix by adopting this theoretical explanation. In practice the error statistics are empirically determined by exploiting their dependence to *a priori* information. As a matter of fact, the background realization $\boldsymbol{x_b}$ is commonly assumed to be equal to the average of the true state $\overline{\boldsymbol{x_t}}$, changing Equation (7.17) to:

$$\varepsilon_b = \boldsymbol{x_t} - \overline{\boldsymbol{x_t}} \tag{7.20}$$

This is a legitimate assumption considering that the model can be viewed as an average of the reality.

For ionospheric applications, a notable example of how the background covariance matrix can be built is well explained in the IDA3D paper by *Bust et al.* [2004]. This technique inspired the approach used for ANIDAS, which exploits ΔfoF2 correlation coefficients reported in the work of *Rush* [1976]. The values are displayed as functions of distance, time of the day in Local Magnetic Time (LMT), season and horizontal directions (east-west and south-north). The algorithm in ANIDAS responsible for building the **B** matrix, initially creates, for a given time of the day and year (values are interpolated between bins of data to avoid abrupt changes), a dense reference matrix of correlation coefficients for the two horizontal directions. The rows of this matrix are populated with coefficients that refer to the east-west direction, the columns to the south-north

one. By using an elliptical function, intermediate correlation values are extrapolated in order to fill the rest of the reference matrix. Figure 18 shows an example of reference matrix of ΔfoF2 correlation coefficients for the 29th of December at 1900 UT.

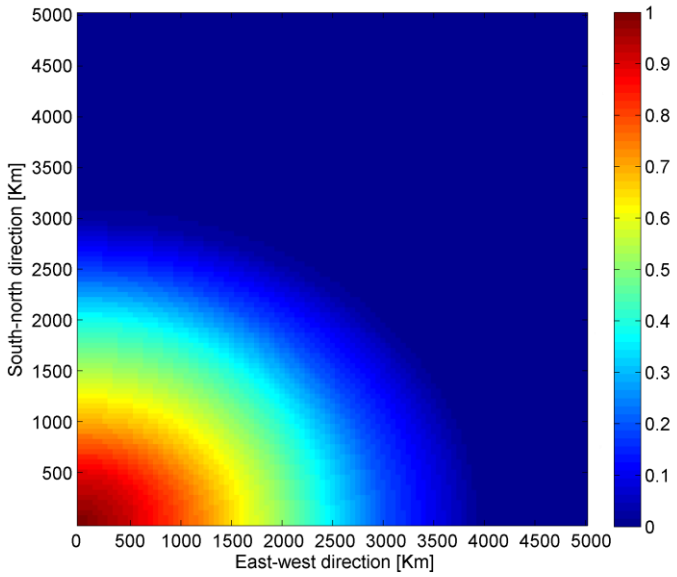

Figure 18. Dense reference matrix of ΔfoF2 correlation coefficient values built for the 29th December at 1900 UT.

As expected, at distance zero the coefficients are close to 1 while at greater distances the correlation fades to zero. Furthermore, it is possible to see that the relationship between coefficients and distances is different according to the direction. Once the reference correlation matrix has been built, it is applied to a given three dimensional grid of positions: from the input grid, the lattice of points that lays at 300 km is considered; for each point of this lattice, a set of coordinates of the all the other points of the lattice is calculated. This means that, if for example the input grid has 6 positions latitude and 11 in longitude, the result of the calculation would be a 66x66 matrix where each element contains a couple of coordinates. The latter are then interpolated with the reference correlation matrix. The final result for a rotated grid centred at 38° latitude and 263° longitude with 6 positions (4° apart) for the latitude and 11 positions (4° apart) for the longitude is shown in Figure 19.

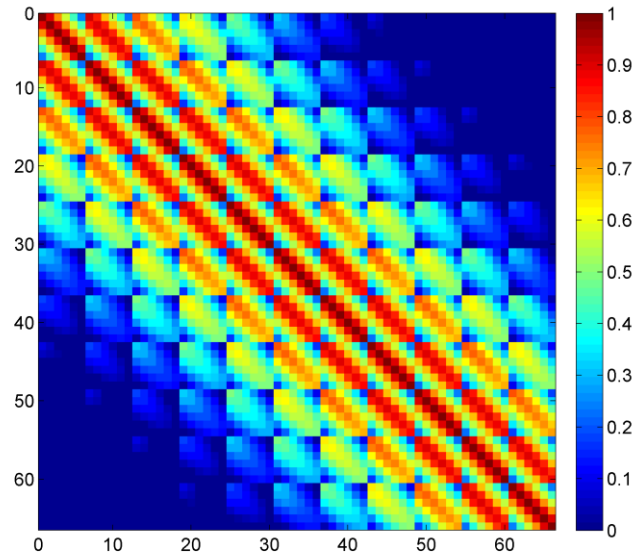

Figure 19. Correlation matrix for a bi-dimensional grid 6x11 (In particular, for a rotated grid that lays at 300 km of altitude, centred in 38˚ latitude and 263˚ longitude with 6 positions (4˚ apart) for the latitude and 11 positions (4˚ apart) for the longitude)

Figure 19 shows the graphical representation of a specific correlation matrix where it is possible to notice that the diagonal is populated by correlations of values 1. To consider the correlation for the altitude, the matrix needs to be expanded by the number of height points defined in the input grid. The horizontal matrix is then simply replicated along the diagonal. Figure 20 shows an example of a complete correlation matrix for 4 height points.

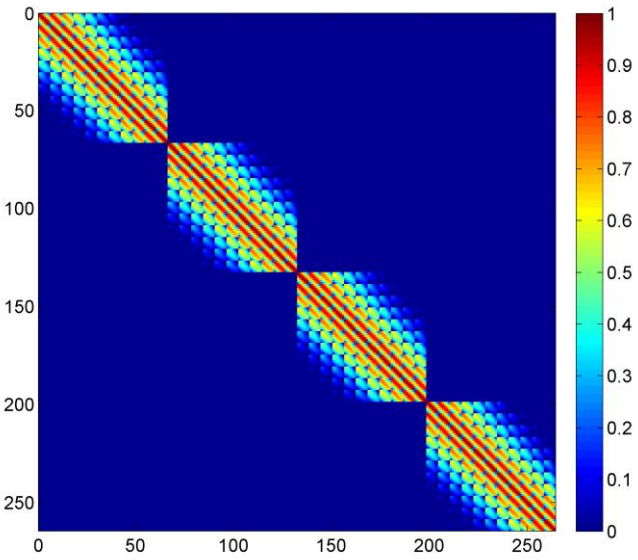

Figure 20. Correlation matrix for a tri-dimensional grid 6x11x4 (In particular, for a rotated grid with 4 points of altitude, centred in 38˚ latitude and 263˚ longitude with 6 positions (4˚ apart) for the latitude and 11 positions (4˚ apart) for the longitude)

The matrix from Figure 20 is now considering three dimensions and, according to the example followed, its size is now 264x264 (6x11x4). It is relevant to note that the diagonal of 1 is maintained. As mentioned in previous sections, the data used in ANIDAS are collected over a window of time divided in time frames where the central corresponds to the time of the reconstruction. Although the background matrix is not theoretically related to the observation collection, it is important to consider the correlation among time frames. If this is not taken into account, the reconstructions performed in each time frame would be disconnected to the others.

The covariance matrix in ANIDAS is generated by considering the number of time frames and their correlation.

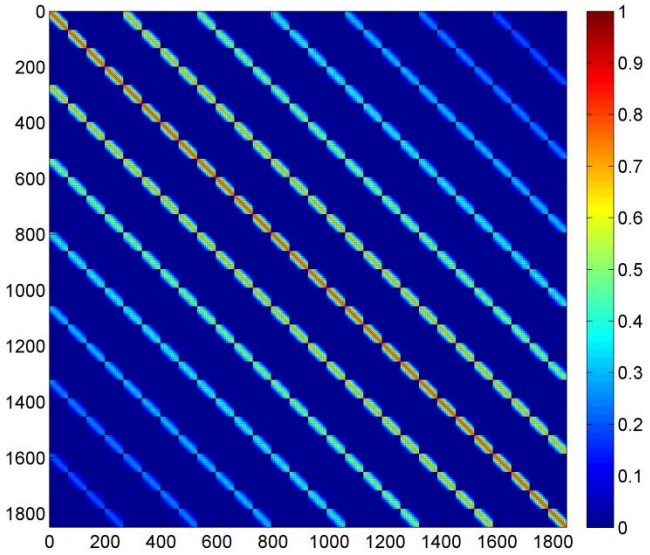

Figure 21. Correlation matrix for a tri-dimensional grid in time 6x11x4x7 (In particular, for a rotated grid with 4 points of altitude, centred in 38° latitude and 263° longitude with 6 positions (4° apart) for the latitude and 11 positions (4° apart) for the long for 7 windows of collection time (10 min apart))

Figure 21 represents the correlation matrix built for a case where 7 time frames are used; its size is therefore 1848x1848 (6x11x4x7). Once again, it is possible to notice that the diagonal is conserved and that the previous matrix 264x264, (shown in Figure 20), is somehow replicated. In Figure 21, the time replicas values fade to zero as they get further from the diagonal. The magnitude of the fading is definable by the user. The structure is produced by combining the previous matrix 264x264 of Figure 20 to a Toeplitz matrix $\mathbf{T}$ where each element $\mathbf{T}_{i,j}$ is defined as follows:

$$\mathbf{T}_{i,j} = {t_{cc}}^{abs(i-j)} \tag{7.21}$$

Where the row index i and the column index j go from 1 to the number of the considered time frames, and t_{cc} is the time correlation coefficient. t_{cc} gives the magnitude with which the correlation fades in time, in the presented example (Figure 21) t_{cc} is set 0.8 for a ten minutes time

frame. In order to have an idea of the trend of correlation fading, the values of the first row of **T** used in the example are plotted in Figure 22.

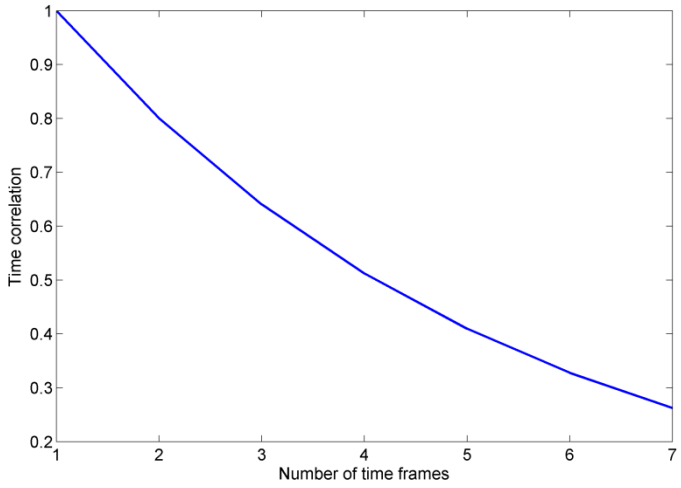

Figure 22. Time correlation exponential decreasing in time by setting time correlation coefficient at 0.8 for 10 minutes time frames.

According to these settings, the Toeplitz matrix **T** looks like the following:

$$\mathbf{T} = \begin{bmatrix} 1.0000 & 0.8000 & 0.6400 & 0.5200 & 0.4096 & 0.3277 & 0.2621 \\ 0.8000 & 1.0000 & 0.8000 & 0.6400 & 0.5200 & 0.4096 & 0.3277 \\ 0.6400 & 0.8000 & 1.0000 & 0.8000 & 0.6400 & 0.5200 & 0.4096 \\ 0.5200 & 0.6400 & 0.8000 & 1.0000 & 0.8000 & 0.6400 & 0.5200 \\ 0.4096 & 0.5200 & 0.6400 & 0.8000 & 1.0000 & 0.8000 & 0.6400 \\ 0.3277 & 0.4096 & 0.5200 & 0.6400 & 0.8000 & 1.0000 & 0.8000 \\ 0.2621 & 0.3277 & 0.4096 & 0.5200 & 0.6400 & 0.8000 & 1.0000 \end{bmatrix} \quad (7.22)$$

Once combined by the Toeplitz matrix, the correlation matrix here is used as a background covariance matrix (the **B** term in Equations (7.8) and (7.15)). According to the relation shown in Equation (7.19) the correlation matrix should be multiplied by $\sqrt{\mathrm{var}(e_i)\mathrm{var}(e_j)}$ in order to obtain a proper covariance matrix. In this work, the parameter α (Eq. (7.8)) acts like the missing term. Because the covariance matrix is inverted (Eq. (7.15)) and by assuming that all the variances are equal, it is possible to assign a mathematical meaning to the parameter α:

$$\alpha = \frac{1}{\sigma^2} \quad (7.23)$$

Where σ^2 is the generic variance valid for all the system. The parameter α is defined by the user and its role can be compared to the parameter λ in MIDAS (Eq. (7.3)). In ANIDAS, by strongly increasing the value of α, the contribution of the background increases (Eq. (7.8)). This translates in an inversion that is more driven by the information given by the background term. On the contrary, by setting the value of α to zero the analysis will rely solely on the observation. For intermediate values, α also regulates the importance of the background covariance matrix **B**, which

in ANIDAS corresponds to the correlation one. As α increases, the operation will respect the correlation provided by **B** even if they contradict the observation; as α decreases, the operation will match more the observations even if they contradict **B**.

Several assumptions related to the production and the usage of the background covariance matrix in ANIDAS can be deduced from the previous explanation. The first one is that the correlation coefficients on which **B** is based are assumed to be valid globally and in every geomagnetic or solar activity. This could be problematic in particularly perturbed events or in peculiar location such as the Polar Regions. However, it should be noted that this research is currently focussed on mid-latitudes and the extension of the techniques to high and low-latitudes is planned as a future extension to this project.

The second assumption is that horizontal correlations can be described by means of an elliptic function (Figure 18). The third one is that horizontal correlations are maintained along the entire profile. ANIDAS uses a tri-dimensional grid of points that is defined by latitude, longitude and altitude. Therefore the distance in km between two points at a certain altitude is different from the respective points at a different altitude. For example, two points at the bottom of the grid are much closer with respect to the respective that lie at the top height. Since at lower altitudes the ionosphere is more structured, one expects that the correlation distances would be shorter. On the other hand, because the ionosphere can be assumed smoother at higher altitudes the correlation distances should be longer. According to this, the third assumption, which implies that the correlation between the two points is the same even if they lay at different heights, is considered acceptable. The fourth assumption is that there is no correlation along the vertical as it is already taken into account by implementing vertical basis functions.

7.2.2 Vertical Basis function

In order to overcome the lack of vertical resolution, in MIDAS, and in ANIDAS, *a priori* information is added. This can be seen as a constraint in the inversion for having more realistic vertical electron density distributions. A generally accepted approach to implement this external information involves the usage of vertical basis functions calculated through a Singular Value Decomposition (SVD) method. To date in MIDAS, they have been empirically extrapolated, for example from a series of pre-modelled profiles by Chapman function or by IRI. The result is a set of Empirical Orthonormal Functions (EOFs) that can be used to control the peak height, the thickness of the profiles and the shape of the lower ionospheric layers. The application of these functions was already introduced in Section 7.1 by Equation (7.4). Another advantage of replacing a whole profile of values $\boldsymbol{n}$ with several coefficients $\boldsymbol{x}$ is that the dimensionality of the inverse problem is drastically reduced, and therefore it can be solved more efficiently. In addition to this, the usage of EOFs is extremely important in order to overcome the sparsity of **H** due to the lack

of data; a single value of **H** intersecting a particular profile is in fact enough in order to represent the entire profile. It is obvious how crucial the role of vertical basis functions is in the solution of the ill-posed problem and in general in the application of the data ingestion scheme.

In ANIDAS, ANIMo is used not only as the background term but also for generating vertical basis functions (Figure 17). In particular, two different basis function generation methods were developed and compared. In the first method (for convenience it will be referred to as method A or A), each profile from the same model realization used as background in the DA scheme is adopted as the vertical basis functions. In other words, these functions are not computed by any orthogonal decomposition methods. For each latitude-longitude location on the input spatial grid, the respective vertical profile of the model realization is normalized and directly used to characterize its corresponding profile when solving the problem. This means that the inversion is solved for one coefficient (per each latitude-longitude location) that coincides to a simple scaling factor of a selected modelled profile. The second approach (method B or B) applies the SVD method to an ensemble of modelled profiles. The latter is produced by running ANIMo over a defined location and time and by setting model parameters to vary over given ranges. In this case the number of the coefficients depends on the number of basis functions which can be determined by the user at the beginning of the process.

7.3 Now-casting and forecasting with ANIDAS

The procedure described so far actually refers to a single cycle of ANIDAS (Figure 17). However, importantly, the system can be used in an iterative manner in order to generate more accurate analysis for each time step (Figure 23).

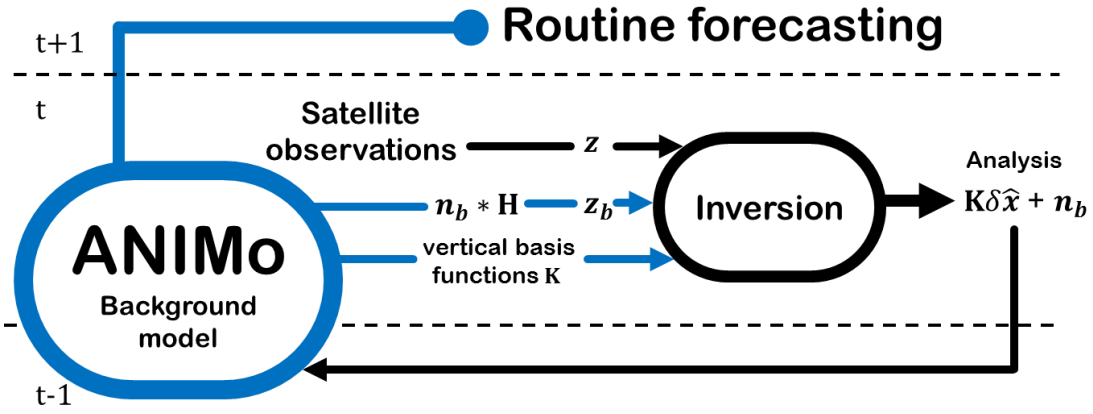

Figure 23. Simplified diagram of the principal components of ANIDAS scheme used in an iterative manner. The letter t indicates time step.

The key point to note is the re-initialisation of the system which corresponds to the initialisation of ANIMo. This is done by feeding ANIMo at the current time step with the analysis produced at

the previous one. ANIMo can then produce a new background term for the current time and therefore trigger a new cycle. Because ANIMo is based on the solution of the O^+ continuity equation (Eq. (5.11)) it has to be initialised with an ionospheric reconstruction in terms of O^+. Given that method A is used, it is possible to retrieve all the coefficients that were applied to the vertical basis function to obtain the final analysis. These coefficients can be seen as scaling factors of the background model. As described in Chapter 5 ANIMo is able to provide O^+, O_2^+, and NO^+ density values which are assumed to give electron density when summed. Hence, by saving the realization for the background state also in terms of O^+, the latter can be scaled in order to find the analysis in terms of O^+. The O^+ analysis is then ready to be used to initialize ANIMo. In the next time step the background model will be more accurate as it corresponds to the modelled evolution of the previous analysis. At any time the system can perform forecasting by using the initialization to run ANIMo forward in time (routine forecasting in Figure 23).

7.3.1 Theoretical comparison between methods A and B and their usage in now-casting mode of ANIDAS

Method A, where the vertical constraint corresponds to the respective profile from the background realization, works better when the model is close to reality. Because method A uses a scaled version of the background, the resulting analysis profile can assume inaccurate profile's shapes when the model is very far from the true state. In contrast, method B is more flexible as it uses ensembles of possible profiles generated by tuning model input parameters, hence can cope better when the background is not very accurate. In addition, the way of generating the ensemble compensates for the uncertainty in setting the model forcing parameter. The disadvantage of method B is that it must rely on data to estimate the profile even if they are generated by a physics-based model; by contrast method A relies solely upon the physical model itself. Approach B can be therefore problematic when there is no data coverage. In these cases the inversion is likely to be not reliable and therefore method B could generate misleading profile shapes. Method A also requests less computational load as it adopts the same model realization used in the DA approach as background information. In method B, the model is asked to create a series of different reconstructions which require multiple runs. This is inconvenient when ANIDAS is used in now-casting. In addition to the computational effort issue, method B is problematic in model initialization phases. Because ANIMo is based on the continuity equation of the monoatomic oxygen ion, the density of the latter is required to initialise the model in the following time step. As mentioned in the previous section, this means that it is necessary to extract the O^+ density information from the resulting DA analysis which is expressed in terms of electron density. If method A is adopted, this can be easily done by scaling ion density profiles generated by ANIMo according to the electron density analysis. The same thing cannot be performed as easily when

using method B because the analysis does not correspond to a scaled version of the background state.

Summary

The general aim of the presented project was to use a physics based model to support ionospheric tomography. In particular, the idea was to use the ionospheric model ANIMo to overcome the issue of the lack of data and poor vertical resolution. The first problem is due to an uneven distribution of GPS ground-based receivers on the Earth surface. In addition, there are locations where there is no data coverage, such as in the middle of the ocean. The second issue is related to the receivers-satellite arrangements, which cannot provide a surrounding scan geometry. For these specific reasons ANIMo was developed and implemented as a background model into MIDAS through a DA scheme. The scheme was created by modifying MIDAS algorithms. In this case, MIDAS cost function takes into account the modelled information generated by ANIMo (background state). In contrast with MIDAS, the DA scheme is working with residual, in other words, the minimization of the new cost function is solved for a gain value. This gain represents the correction that has to be added to the background state to have the final analysis. Working with residual raised an issue related to the handling of the GPS measurements offsets with residual values. The presented DA scheme uses MIDAS calibration that is entangled with the calculation of the final solution. It was demonstrated that MIDAS can work with residual values. An important aspect of the presented DA scheme is also the usage of a background covariance matrix. This is built by a time-dependent function that extrapolates horizontal correlation distances and coefficients from ionosonde historical data series. ANIMo is used to not only provide the background state but also for generating vertical basis functions. Their purpose is to constrain the reconstruction of the vertical electron density profiles during the inversion. This allows it to shape each profile by following the physics information of the model. Two different approaches to generate vertical basis function were developed in this project. One extrapolates them directly from the background state; the other performs a SVD over an ensemble of profiles produced by ANIMo by tuning its forcing parameters in order to simulate different possible states of the ionosphere.

Since the implementation of ANIMo involves a DA method, the presented set up can be used iteratively for performing ionospheric now-casting and eventually forecasting. This is possible by re-initializing ANIMo at every current time step of scheme with the analysis from the previous time step. At a given time, ANIMo can be initialized to run forward in the future for producing forecasting. The presented set-up in its entirety is called ANIDAS and it represents the evolution of MIDAS towards a more physic driven imaging tool and forecasting system.

Chapter 8 ANIDAS results

Abstract

In this chapter the DA suite called ANIDAS is assessed. For a chosen case study, simulated data coverage scenarios determine if the scheme is able to deal with changing densities of observations across a region. Preliminary results, where the model ANIMo is used to provide *a priori* information, demonstrate that ANIDAS can efficiently handle different levels of data coverage at the same time. The comparison against ground instruments, such as ionosondes in various locations of the studied area, proved a validation of the vertical profile up to the peak height. The last aspect is indeed critical in ionospheric tomography application. Because ANIDAS is a data ingestion technique, it can be used iteratively over time by re-initializing the background model ANIMo. Ionospheric forecasting capabilities of the whole setup are also tested. The validation shows that although further investigations and refinements should continue, the outcomes of ANIDAS in now-casting and forecasting mode are very encouraging.

Introduction

The validation of a scheme like ANIDAS involves several testing phases that include comparisons between its results and the real measurements. According to the general objectives of this project, ANIDAS is supposed to perform various functions. One of the functions is to improve ionospheric mapping by supporting tomographic imaging. In particular, issues like handling the lack of data coverage and poor vertical resolution are crucial. Powered by the physics-based model ANIMo, ANIDAS has also the potential to be used in order to forecast the ionosphere.

Testing these capabilities requires the setup of an ad hoc experiment. The general functionality of the scheme is checked by observing how it deals with good data coverage and very poor coverage across an extended region at the same time. This is a situation that very often occurs when using CIT. The examination of the final outcomes and their comparison against real data states the level of accuracy that can be achieved. Performing this assessment from a vertical view point can demonstrate whether the scheme can actually improve the vertical resolution. This type of validation can be expanded over time to see if not only spatial but also temporal features can be recognized. This is valid also for the forecasting capabilities if, at a certain time, ANIDAS analysis is used as initialization of ANIMo and the latter is run forward in time to predict the behaviour of the ionosphere.

The present chapter collates the preliminary results of ANIDAS used in now-casting and forecasting modes. In particular Section 8.1 reports the details of the case study used as validation experiment. Section 8.2 is dedicated to ANIDAS in now-casting and it is divided in two Subsections: 8.2.1 and 8.2.2, where results of the first time step and consecutive time steps are displayed. Section 8.3 is focussed on the forecasting capabilities of the method.

8.1 Case study

For the case study, an ionospheric volume over the US area was selected. Figure 24 shows the projection on the Earth surface of the selected grid of points. The altitude range of the three-dimensional grid goes from 80 to 600 km altitude in 10 km steps, the latitude and longitude intervals are both set at 4°. To maintain a certain consistency and because of a good data availability, viz. ionosonde measurements in different locations in the selected area, the case study corresponds to the one that was used for ANIMo validation in the winter season. This is also convenient as the ANIDAS reconstruction procedure requires large computing effort. The experiment aims to reproduce the ionosphere on the 29th of December 2011 at 1900 UT, a day with medium solar intensity (*F10.7*: 142.3) and with an unperturbed geomagnetic condition (A_p: 9). This information corresponds to the input specifications used to run ANIMo as background in ANIDAS. In addition to that, the vertical velocity adjustment and the top boundary condition are set to default value.

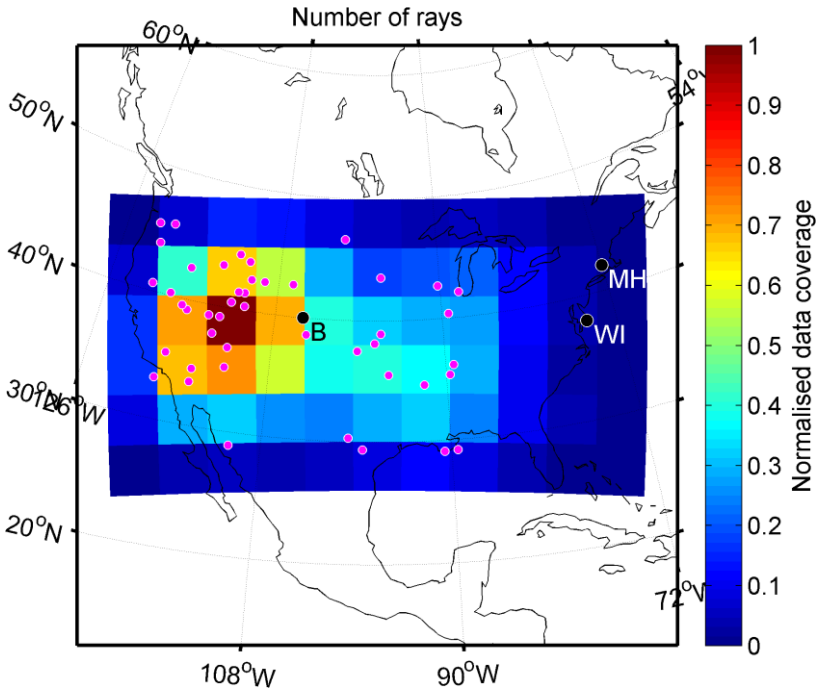

Figure 24. Ground station selection and data coverage map for the 'cold start' case study (29th of December 2011 at 1900 UT)

Magenta dots in Figure 24 represent the locations of the used ground-based GPS receivers. The colour of each 4˚x4˚ tile is associated with the number of satellite-receiver rays and indicates the amount of data coverage. In particular, Figure 24 shows the number of rays normalised to 1 that have been taken into account in the process; the most populated tile has 5671 rays. As aforementioned in Section 7.1, these observations are collected at a rate of 10 minutes within multiple time windows. For this experiment, 7 windows, 3 before and 3 after the selected time (1900 UT), each 10 minutes long were considered. From Figure 24, it is possible to notice that the ground-based receivers are homogenously distributed in the central-west part of the grid and completely absent in the eastern part. The eastern receivers were intentionally removed to test the capability of ANIDAS to deal with severe lack of data like in the middle of the ocean. The advantage is that the selected 'low data coverage' area contains several ground based instruments for monitoring the ionosphere, especially for vertical profile, which are used here to validate the accuracy of the presented methodology. In particular, data from the ionosonde situated in NASA Wallops Islands Flight Facility in Virginia (Lat. 37.5˚, Lon. 284.7˚, highlighted by the black dot followed by WI in Figure 24 and in all the following maps) are considered. As shown in Figure 24, the data coverage of the Wallops Island facility is particularly low but not absent. The site of the Massachusetts Institute of Technology (MIT) Haystack Observatory at Millstone Hill in Massachusetts (Lat. 42.6˚, Lon. 288.5˚, black dot followed by MH) can be almost considered in 'no data coverage' zone. The data from the observatory ISR and ionosonde were also collected and compared with ANIDAS scheme outcome. The ionosonde in the National Oceanic and Atmospheric Administration (NOAA) facility of Boulder in Colorado (Lat. 40.0˚ Lon. 254.7˚, black dot labelled by B) is taken into account as well at the first time step. In contrast with the first two sites, the latter is located in an area with good data coverage.

8.2 ANIDAS Now-casting

8.2.1 The 'cold start'

A realization of the ionospheric model ANIMo was used as *a priori* information in the ANIDAS scheme (described in Section 7.1) by following the input settings reported in the previous section (8.1). Figure 25 shows the realization plotted over the area selected for this case study. The colour of each tile is associated with TEC values calculated from the simulation of ANIMo.

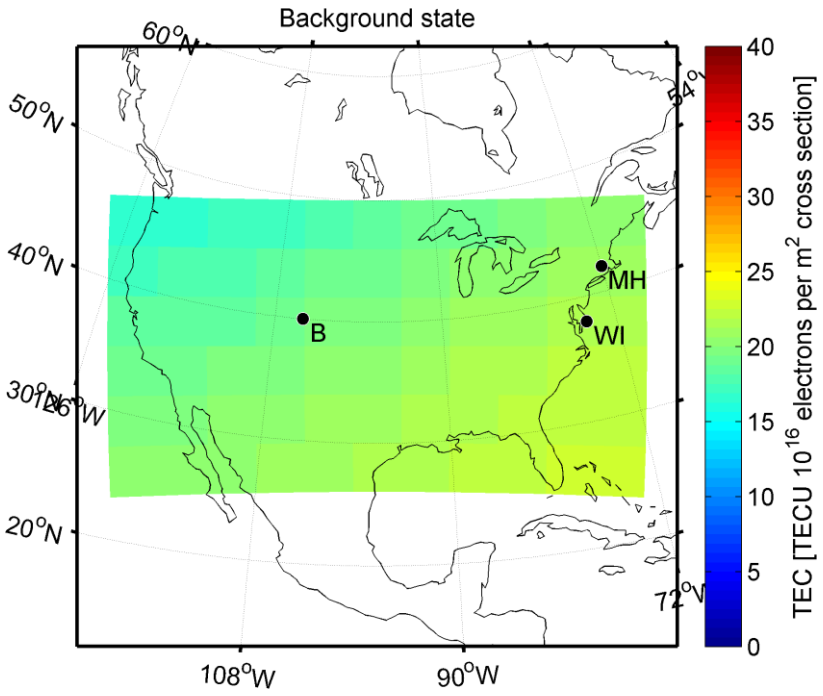

Figure 25. TEC map of the background state used in the 'cold start' case study (29th of December 2011 at 1900 UT)

By using the ANIMo realization as background information in ANIDAS, it was possible to calculate the analysis increment (also known as gain) as a result of the cost function minimization (Eq. (7.8)). Figure 26 shows TEC values extrapolated by the gains of two distinct approaches. The first graph in the picture refers to the outcomes determined by using method A, the second one by using method B. As mentioned in Subsection 7.2.2, these are two different approaches to implement vertical basis functions in the inversion process. Method A applies ANIMo background realization as vertical basis function, to be more specific: for each latitude longitude location, the electron density profile in the inversion is constrained by the respective background one. In contrast, method B (EOF-type approach) generates its basis functions by applying SVD decomposition to an ensemble of electron density profiles produced by ANIMo by tuning its input parameters. For the 'cold' start experiment α was set to 0.2 for method A and to 0.25 for method B.

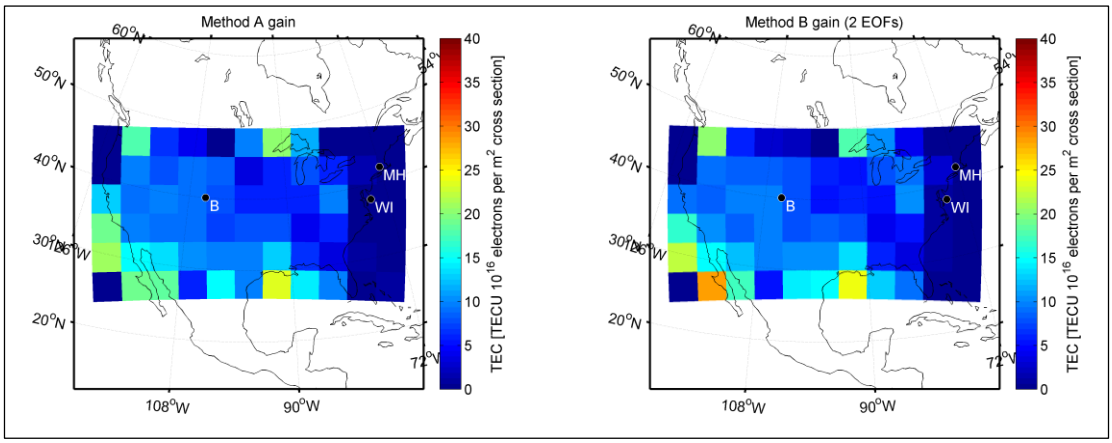

Figure 26. TEC maps of the gain values calculated by ANIDAS through method A (on the left) and method B (on the right) for the 'cold start' case study (29th of December 2011 at 1900 UT)

The two plots in Figure 26 are almost identical where the lack of data is more important across the east coast (see Figure 24). This is expected as the observations in those areas are missing and so introduce very little or no information to the inversion process so the contribution is almost completely provided by the background. Where the data coverage is more significant, it is possible to notice very small differences between the gain TEC values of the two methods, in particular the B gain looks slightly higher. Section 7.1 explained that to estimate the analysis, it is necessary to add the increment onto the background state. The TEC values of the analysis obtained by the two methods are plotted in Figure 27.

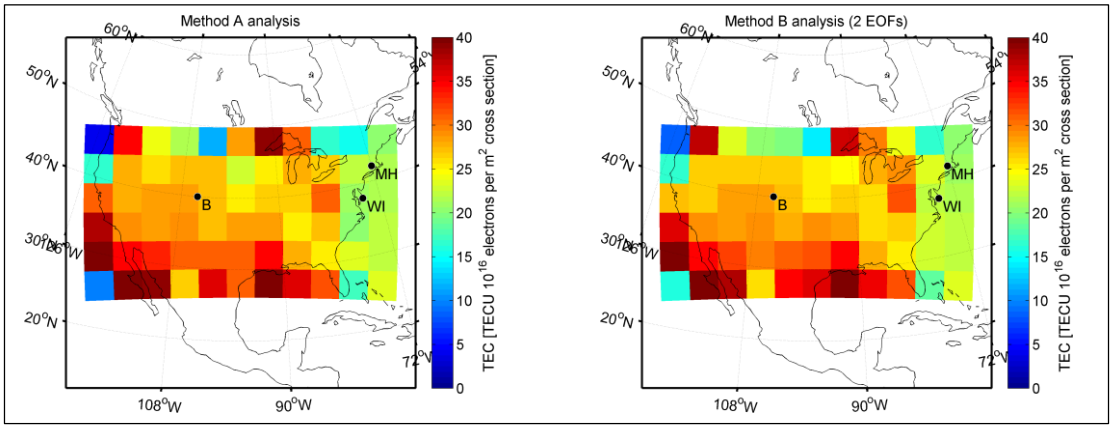

Figure 27. TEC maps of the analysis values calculated by ANIDAS through method A (on the left) and method B (on the right) for the 'cold start' case study (29th of December 2011 at 1900 UT)

As expected, very similar increment values (Figure 26) generated very similar analysis (Figure 27). In addition to this, it is possible to notice noisy values at the edges of the analysis TEC map from both methods. This is believed to be a border effect that can be caused by several reasons. The first is associated to the fact that the satellite ray-paths of the ground-stations that lay on the edges might be truncated. If this is the case, the algorithm automatically discards the partial observation. This translates into a frame of low data coverage close to areas with sufficient data coverage. In a 'cold start' ANIDAS struggles to find a good solution for these perimeter points. Secondly, the

background covariance matrix is lacking sensible boundary conditions for the moment. This means that the electron density values at the edges can only partially 'benefit' from the additional information of the neighbours voxels. By focussing on the inner area, Figure 27 shows that, with small differences in the two methods, ANIDAS is able to smoothly combine background and observation contributions. Figure 41 in Appendix A, reports some reconstruction of the same case study (intentional data gap inclusive) performed by the standard version of MIDAS. As said in Section 7.1, MIDAS relies almost completely on GPS measurements to generate its results; hence it is fair to consider it as a representation of the observation contribution in ANIDAS. Comparing MIDAS TEC map in Figure 41 (for example the one where $\lambda = 1$) and ANIMo TEC map in Figure 25, it is evident that the two are quite different: MIDAS TEC values are much higher. Considering now also ANIDAS TEC maps in Figure 27, it can be seen that ANIDAS relies more on the observations where the data coverage is abundant and more on ANIMo background state where data is lacking. This is confirmed by observing vertical electron density profiles over Boulder (Co) in Figure 28, Millstone Hill (Ma) in Figure 29, and over Wallops Island (Va) in Figure 30 and by observing the performances statistics reported in Table 16, Table 17 and Table 18 respectively.

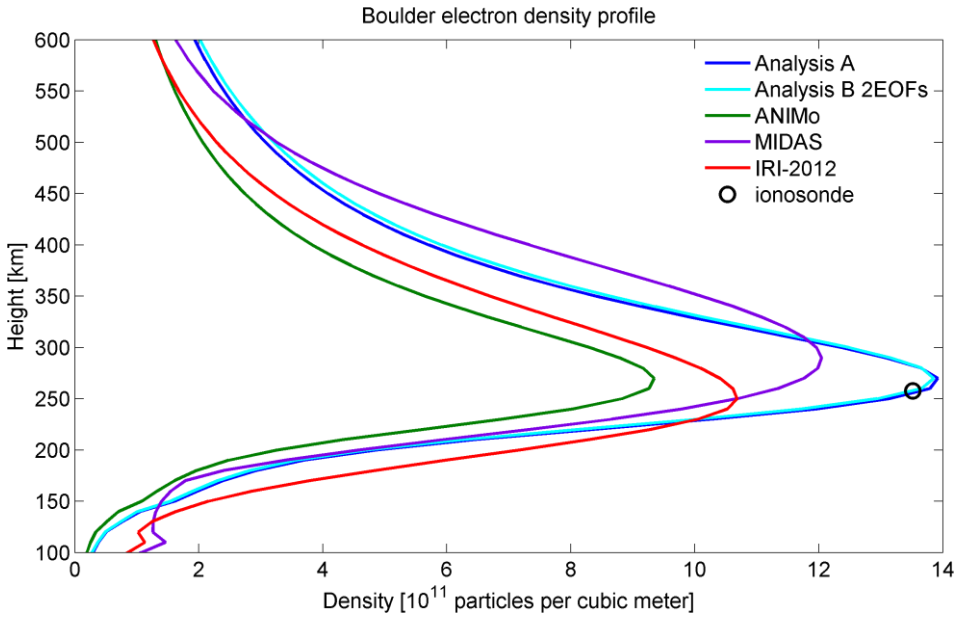

Figure 28. Electron density vertical profiles over Boulder from different sources: ANIDAS method A in blue, ANIDAS method B in cyan, ANIMo in green, MIDAS in purple, IRI-2012 in red and the ionosonde measurement is shown as a black circle. This refers to the 'cold start' case study (29th of December 2011 at 1200 UT (1900 UT))

Figure 28 shows the electron density profile over Boulder digisonde in Colorado provided by different sources. The solid blue and cyan lines correspond to the profiles extrapolated by the analysis of ANIDAS by implementing method A and B respectively. ANIMo simulation is shown in green, MIDAS in purple, IRI-2012 in red and the black circle refers to the ionosonde density at the peak vs peak height measurement. Table 16 reports absolute difference between the NmF2 and hmF2 extrapolated from the different methods and those measured by the ionosonde.

Boulder ('cold start' case study, 29th of December 2011 at 1200 UT (1900 UT))						
NmF2 [10^{11} m^{-3}]		**Method A**	**Method B**	**ANIMo**	**IRI-2012**	**MIDAS**
	Absolute difference	0.405	0.341	4.168	2.828	1.467
hmF2 [km]		**Method A**	**Method B**	**ANIMo**	**IRI-2012**	**MIDAS**
	Absolute difference	12.300	12.300	12.300	7.700	32.300

Table 16. Boulder location statistics ('cold start' case study, 29th of December 2011 at 1200UT (1900 UT)). The table reports the absolute difference between each method and the ionosonde.

The comparison between IRI-2012 and the ionosonde tells that at this particular location and time the ionosphere was higher in density than the model IRI-2012 indicated. The green profile reveals that ANIMo is not able to model accurately this specific unexpected enhancement. In contrast MIDAS (in purple) is sensitive to it, aided by the very good data coverage of the area. This means that among the MIDAS and ANIMo TEC maps (Figure 41 and Figure 25 respectively), the first is definitively more accurate where there is data coverage. The vertical profile from MIDAS is, in this reconstruction, driven by 2 EOFs produced by IRI-2012, which provide a plausible shape but not very accurate. ANIDAS profiles (in blue and cyan) from both methods seem to have a more accurate match with the ionosonde measurement (see Table 16). Furthermore, method A and B profiles are extremely similar and both struggle in reproducing structures at the bottom of the profile.

The key point to note from this validation is that the ANIDAS method based upon the model ANIMo is a more accurate representation of the peak height and density than either model alone or MIDAS alone.

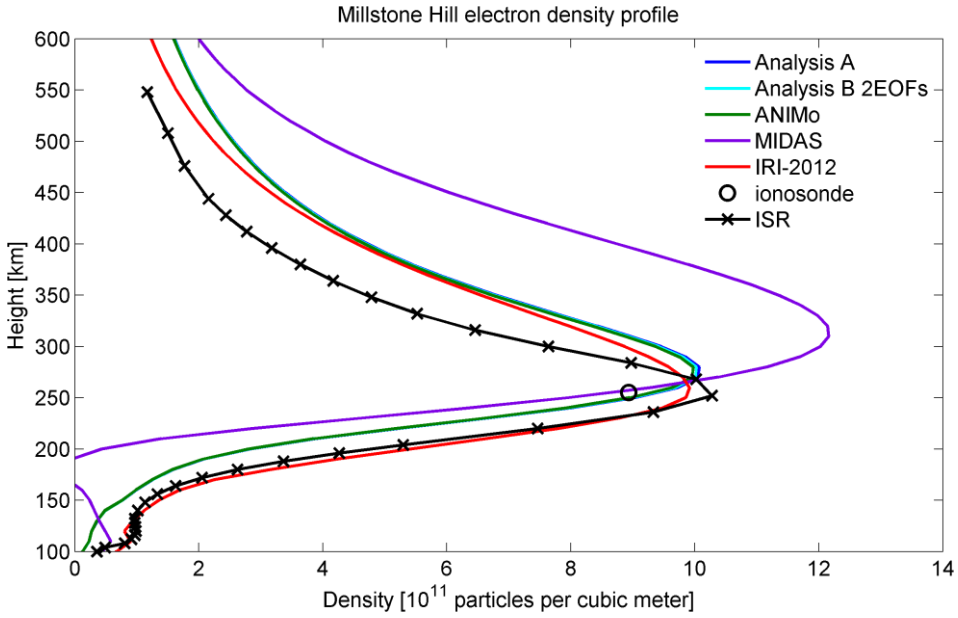

Figure 29. Electron density vertical profiles over Millstone Hill from different sources: ANIDAS method A in blue, ANIDAS method B in cyan, ANIMo in green, MIDAS in purple, IRI-2012 in red, the ionosonde measurement is shown as a black circle and the ISR data are reported by a crossed solid black line where the crosses correspond to the radar ranges. This refers to the 'cold start' case study (29th of December 2011 at 1400 LT (1900 UT)).

In the region of poor data coverage (i.e. the east coast where the receivers were deliberately omitted to simulate the performance over regions of sparse data) the scenario is very different. Figure 29 shows vertical electron density profiles over Millstone Hill where the data coverage is very sparse or there are no data, and Table 17 reports NmF2 and hmF2 comparison statistics.

Millstone Hill ('cold start' case study, 29th of December 2011 at 1400LT (1900 UT))						
NmF2 [10^{11} m^{-3}]		**Method A**	**Method B**	**ANIMo**	**IRI-2012**	**MIDAS**
	Absolute difference	1.140	1.097	1.044	0.985	3.232
hmF2 [km]		**Method A**	**Method B**	**ANIMo**	**IRI-2012**	**MIDAS**
	Absolute difference	24.810	24.810	24.810	4.810	54.810

Table 17. Millstone Hill location statistics ('cold start' case study, 29th of December 2011 at 1400 LT (1900 UT)). The table reports the absolute difference between each method and the ionosonde.

Here ANIMo (green) and both ANIDAS methods (blue for A and cyan for B) are essentially the same. This means that the DA scheme is totally reliant upon the background contribution. In this location, the ionosphere is behaving as expected and this is demonstrated by the similarity of ionosonde (black circle) and ISR measurements (crossed black line, where crosses indicate the range intervals) against the profile from IRI-2012 (red). Apart from the lower structures of the profile, ANIMo is in this case more accurate and so are ANIDAS analyses. On the contrary, MIDAS (in purple) suffers more acutely from the lack of data coverage. The regularization factor (Section 7.1, Eq. (7.3)) tends to extend electron density gradients from highly covered area over non covered area (see Figure 24 and Figure 41). Figure 29 and Table 17 show, as expected, that this solution is not always advisable.

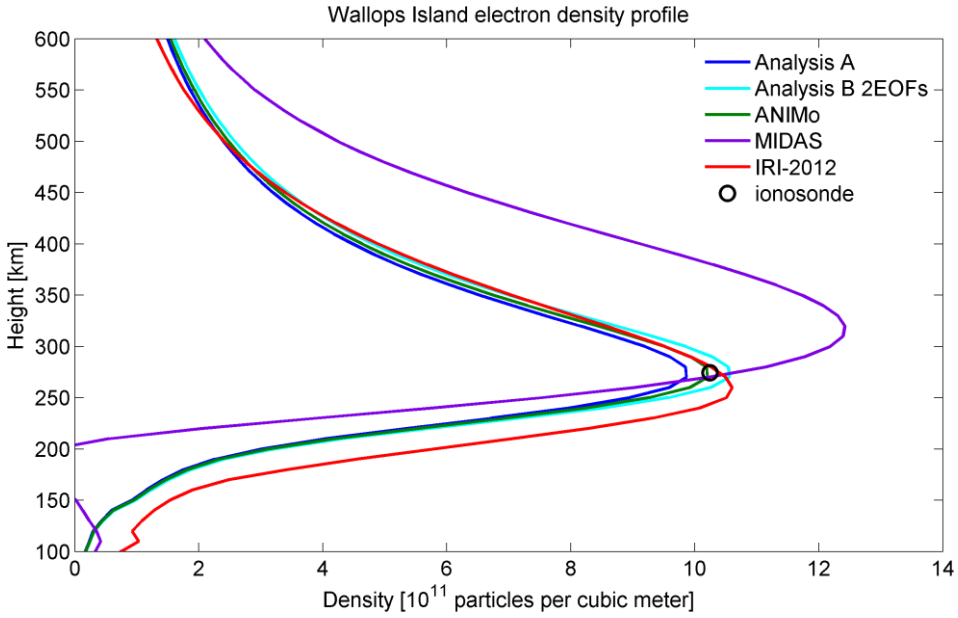

Figure 30. Electron density vertical profiles over Wallops Island from different sources: ANIDAS method A in blue, ANIDAS method B in cyan, ANIMo in green, MIDAS in purple, IRI-2012 in red and the ionosonde measurement is shown as a black circle. This refers to the 'cold start' case study (29th of December 2011 at 1400 LT (1900 UT)).

Similarly for Millstone Hill example, also Wallops Island (Figure 30, Table 18) is located in a very low data coverage area; this is confirmed by the analogous results shown in Figure 29 and Table 17.

Wallops Island ('cold start' case study, 29th of December 2011 at 1400LT (1900 UT))						
NmF2 [10^{11} m^{-3}]		**Method A**	**Method B**	**ANIMo**	**IRI-2012**	**MIDAS**
	Absolute difference	0.375	0.317	0.036	0.361	2.181
hmF2 [km]		**Method A**	**Method B**	**ANIMo**	**IRI-2012**	**MIDAS**
	Absolute difference	4.400	4.400	4.400	14.400	45.600

Table 18. Wallops Island location statistics ('cold start' case study, 29th of December 2011 at 1400 LT (1900 UT)). The table reports the absolute difference between each method and the ionosonde.

MIDAS (in purple) is, also in this case, overestimating the general electron density value and being unreliable at the bottom of the profile. ANIMo (in green) is accurately describing peak altitude and density at the peak. Similarly ANIDAS analyses (in blue and cyan) are close in matching the ionosonde measurement (black circle). Here, it is possible to observe a small difference between method A and B: the latter seems to slightly overestimate while the former seems to underestimate the reference data. ANIMo and the analyses seem, once again, unable to pick the lower layers of the profile.

8.2.2 Background model re-initialisation

The now-casting capabilities of ANIDAS include the possibility to use the current analysis as the initialization of ANIMo. The model is fed with the current analysis output O^+ profile and then run forward one step to produce a new background state for the next analysis. In regions of poor data this can allow the model to successively produce more realistic results by utilising data over multiple time steps. As described in Section 7.3, this function can be used in an iterative manner. It was also highlighted that, because ANIMo accepts only O^+ density values as initialization, it is necessary to use solely method A to produce vertical basis functions. The consecutive hours from the case study presented for the 'cold start', Section 8.2.1, were selected to validate the now-casting capabilities of ANIDAS. In particular, the analysis was saved every 30 minutes from 1900 UT (excluded) to 2100 UT. The same ground-station arrangement is kept for all the time steps hence, despite small differences due to satellites movement, the data coverage is each time very similar to Figure 24. Figure 31, shows the resulting TEC maps. For each 30 minutes analysis, the α parameter in ANIDAS has been modified in order to obtain a reasonable reconstruction. The values of the used α are collected in Table 19.

Values of α parameters used during now-casting				
	1930 UT	**2000 UT**	**2030 UT**	**2100UT**
α	0.1	0.01	0.02	0.03

Table 19. Values of α parameter used in the now-casting experiment.

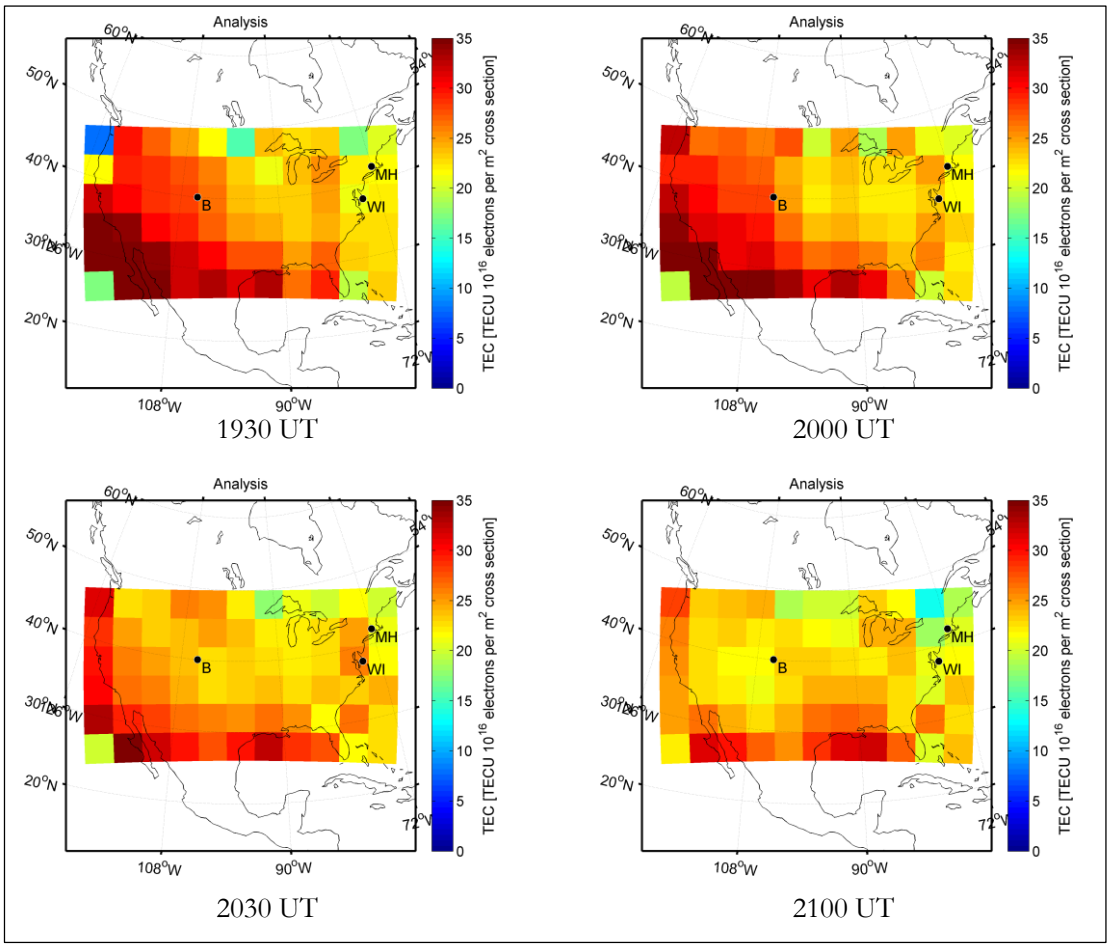

Figure 31. TEC maps of the analysis values calculated by ANIDAS through method A in now-casting mode (29th of December 2011 from 1930 to 2100 UT)

By comparing the TEC maps from Figure 31 with the ones reported in Figure 27 it is possible to notice that the edge effect that was found at the 'cold start' is gradually disappearing. This is probably due to the fact that, thanks to the re-initialization of ANIMo, the background state moves closer to the observation contribution at each time step. As mentioned in Section 7.3, the re-initialization consists of feeding the model with the analysis from the previous time step. The result is hence that the background for the current time step is an evolution of the analysis state of the previous time step. This allows a less traumatic reconstruction at the problematic borders. In general, ANIDAS performs well in combining different contributions and reproduces a reasonable electron density gradient visible in all 4 reconstructions over the selected area. From a vertical point of view, this is confirmed by the comparison of the electron density at the peak and peak height between the analysis and the ionosonde measurements (Figure 32, Figure 33, and Figure 34).

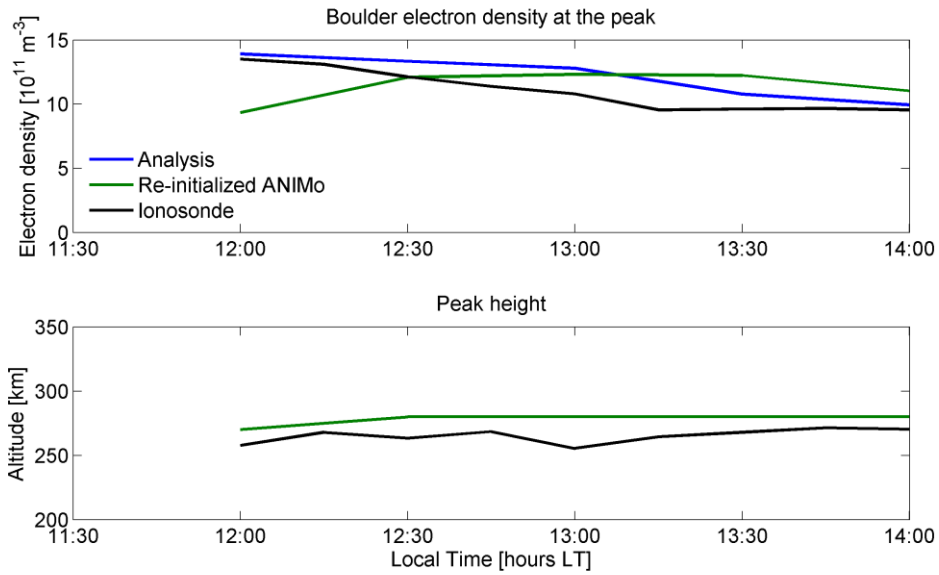

Figure 32. The graphs show, respectively, the comparisons of electron densities at the peak and peak heights produced by ANIDAS method A (blue), by re-initialized ANIMo (green) and measured by ionosonde (black) for the now-casting case study (29th of December 2011 from 1200 to 1400 LT (1900 to 2100 UT)) over Boulder.

The first graph in Figure 32 shows the evolution in time, over the Boulder location, of NmF2 from the anaysis in blue, the background state produced by ANIMo re-initialized in green, and finally the ionosonde data in black. Despite a quasi-systematic overestimation, the analysis seems to follow the trend outlined by the ionosonde measurements. It is interesting to notice that the background state (green line), which at the 'cold start' is significally lower than the the real data, gets 'corrected' by the observations contribution coming from a good data coverage. In Figure 32, the graph beneath shows the evolution in time, over Boulder, of the peak altitude. The line color convention is kept, the reason why the analysis is not visible is due to the fact that, according to method A, the vertical basis functions used in the inversion are a scaled version of the background state and hence the green line overlays the blue line. Once again, the analysis accompanies the ionosonde data with a small and consistent overestimation in the hmF2.

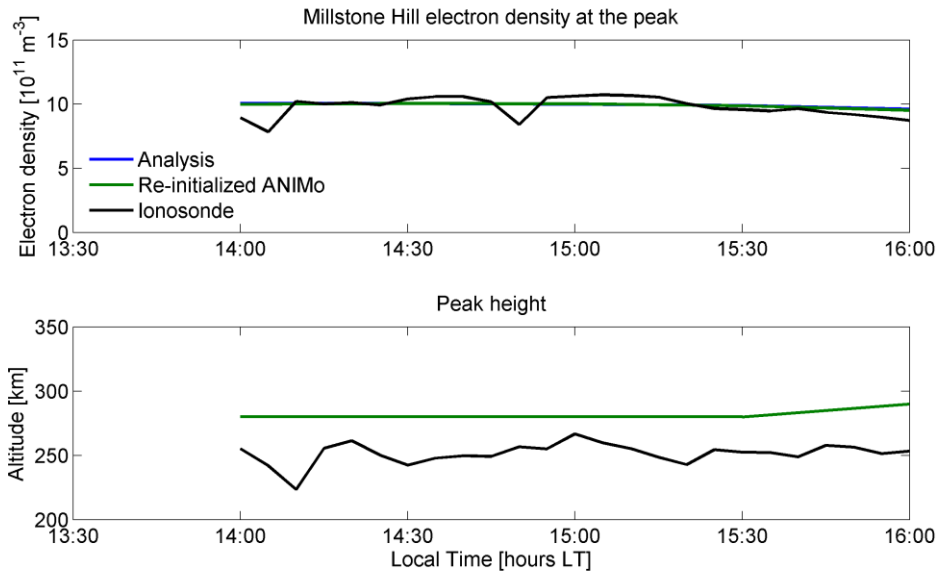

Figure 33. The graphs show, respectively, the comparisons of electron densities at the peak and peak heights produced by ANIDAS method A (blue), by re-initialized ANIMo (green) and measured by ionosonde (black) for the now-casting case study (29th of December 2011 from 1400 to 1600 LT(1900 to 2100 UT)) over Millstone Hill.

Figure 33 refers to Millstone Hill which is located in a low data coverage zone. ANIMo is behaving well in reproducing the elctron density at the peak and this is very important as it basically represents the only contribution in the analysis. With regards to the peak altitude, although an offset is present, the trend is once again well recognized by the DA scheme.

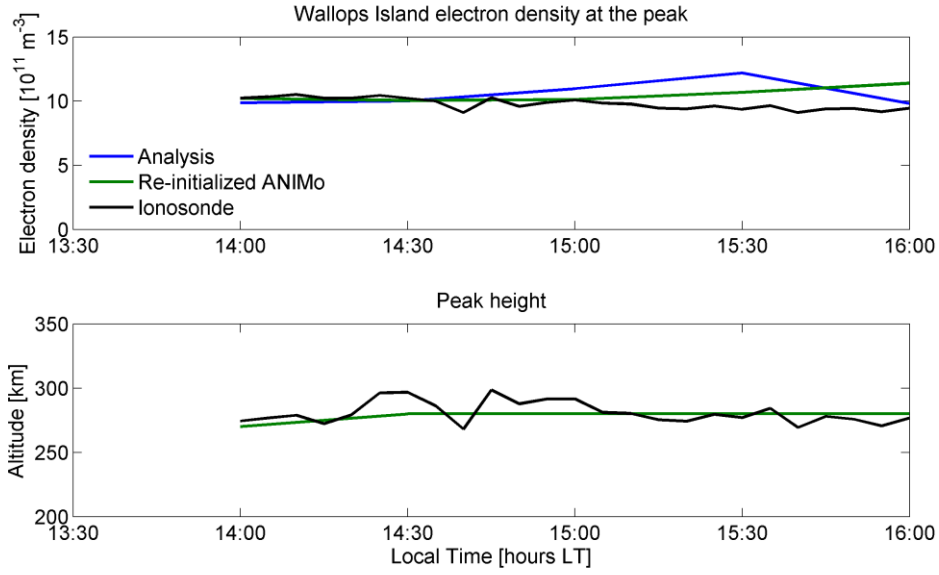

Figure 34. The graphs show, respectively, the comparisons of electron densities at the peak and peak heights produced by ANIDAS method A (blue), by re-initialized ANIMo (green) and measured by ionosonde (black) for the now-casting case study (29th of December 2011 from 1400 to 1600LT (1900 to 2100 UT)) over Wallops Island.

Figure 34 exhibits the same type of graphs explained before but here they refer to Wallops Island location where the data coverage is very low. Regarding the electron density at the peak, there is good agreement between the analysis (blue) and the ionosondes (black) except for the second to last time step. The peak altitude is well modelled by ANIMo and the final analysis benefits from this.

8.3 ANIDAS Forecasting

One of the most interesting applications of ANIDAS is the ability to use it to forecast the ionosphere. In Section 7.3 it was briefly explained that at any point the analysis can be used to initialize ANIMo to run forward in time on its own. In the following examples, the analysis produced by ANIDAS at 2100 UT was fed into ANIMo. The latter was run for several hours and its outcomes were saved every 30 minutes. Figure 35 collates the resulting TEC maps for each time step.

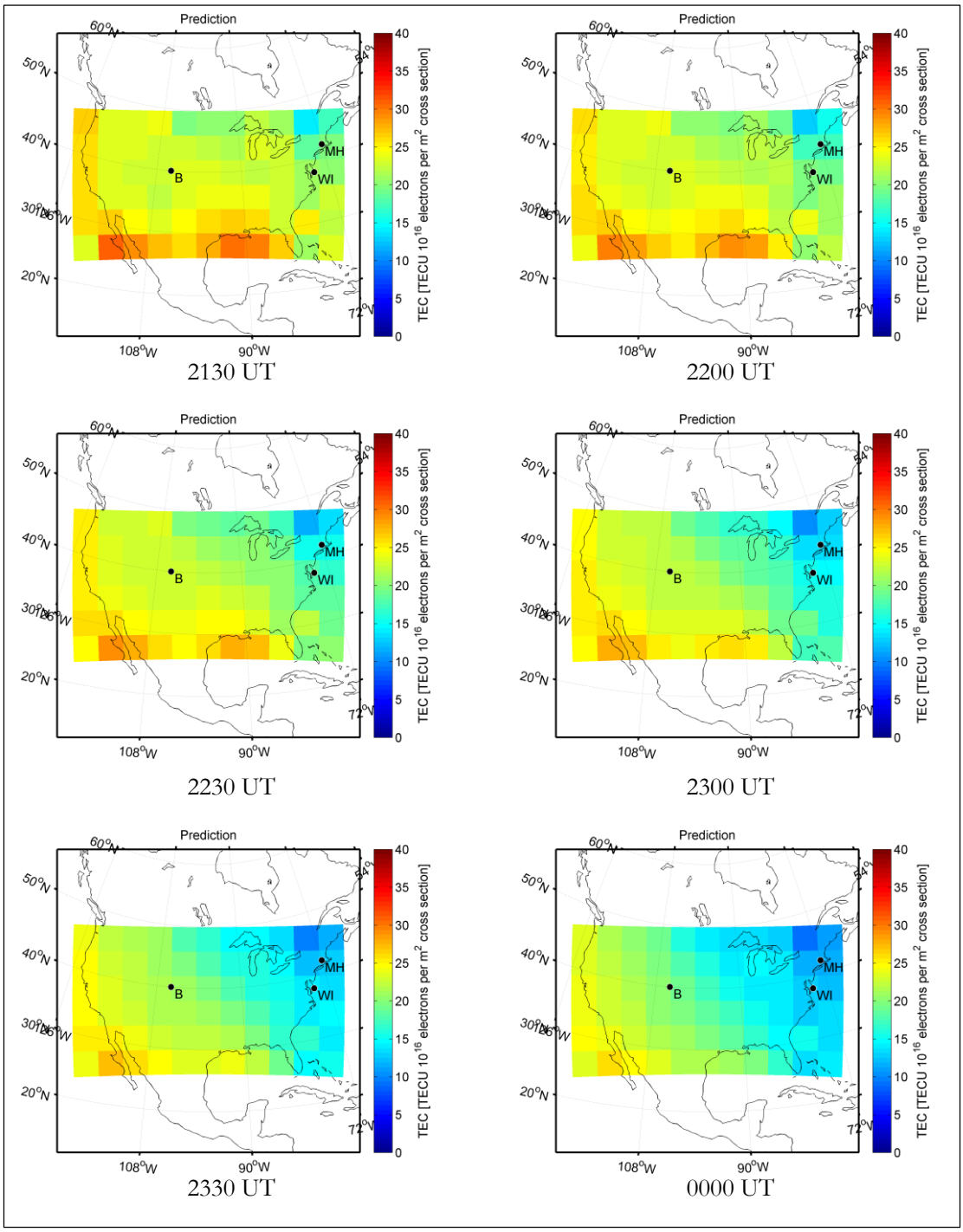

Figure 35. TEC maps calculated by ANIMo in forecasting mode (29th of December 2011 from 2130 to 0000 UT)

Similarly for the TEC maps produced by ANIDAS in now-casting mode (Figure 31), the predicted TEC maps (Figure 35) give a plausible image of the ionosphere. The gradient observed in now-casting is persistent here and moving across the selected area. Once again the vertical prespective for each location can give better insight to the situation and validate the model prediction. The following graphs, in Figure 36, Figure 37, and Figure 38, are an extension in time of the electron density at the peak and altitude peak comparison plots displayed in the previous section in Figure

32, Figure 33 and Figure 34 respectively. In each graph, the highlighted area in light blue corresponds to the 'future' time period.

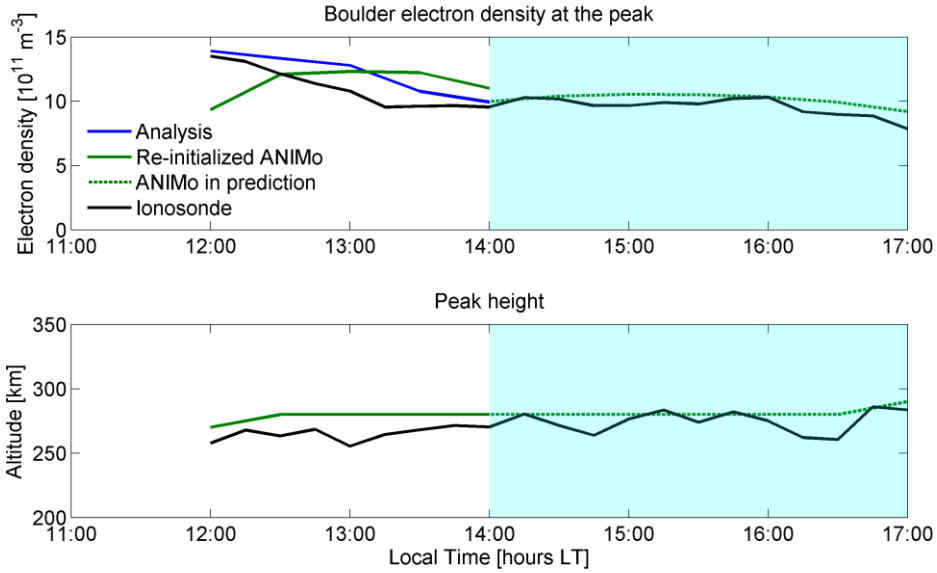

Figure 36. The graphs show, respectively, the comparisons of electron densities at the peak and peak heights produced by ANIDAS method A (blue), by re-initialized ANIMo (green) and measured by ionosonde (black) for the now-casting case study (29th of December 2011 from 1200 to 1400 LT (1900 to 2100 UT)) and by ANIMo (dashed green) in forecasting (29th of December 2011 from 1400 to 1700 LT (2100 to 0000 UT), highlighted in light blue) over Boulder.

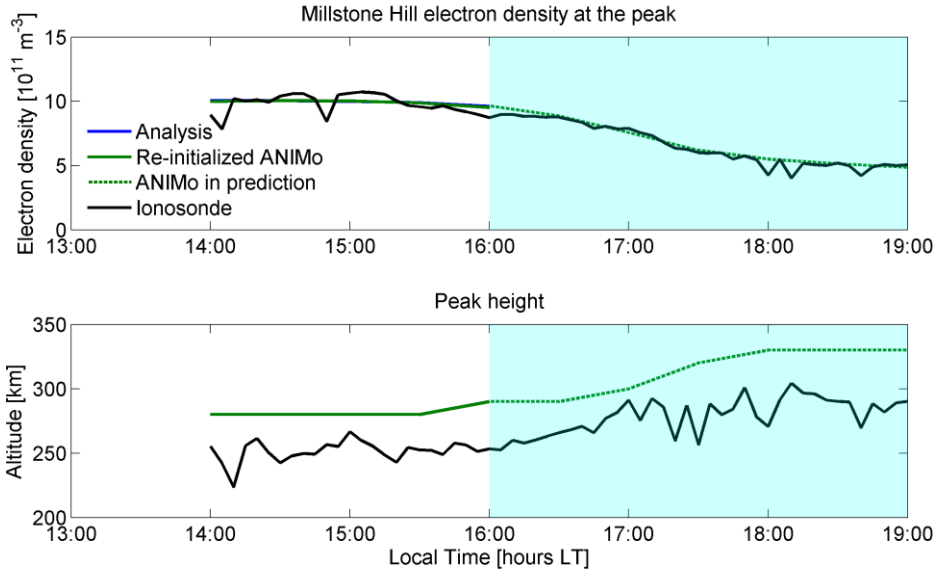

Figure 37. The graphs show, respectively, the comparisons of electron densities at the peak and peak heights produced by ANIDAS method A (blue), by re-initialized ANIMo (green) and measured by ionosonde (black) for the now-casting case study (29th of December 2011 from 1400 to 1600 LT (1900 to 2100 UT)) and by ANIMo (dashed green) in forecasting (29th of December 2011 from 1600 to 1900 LT (2100 to 0000 UT), highlighted in light blue) over Millstone Hill.

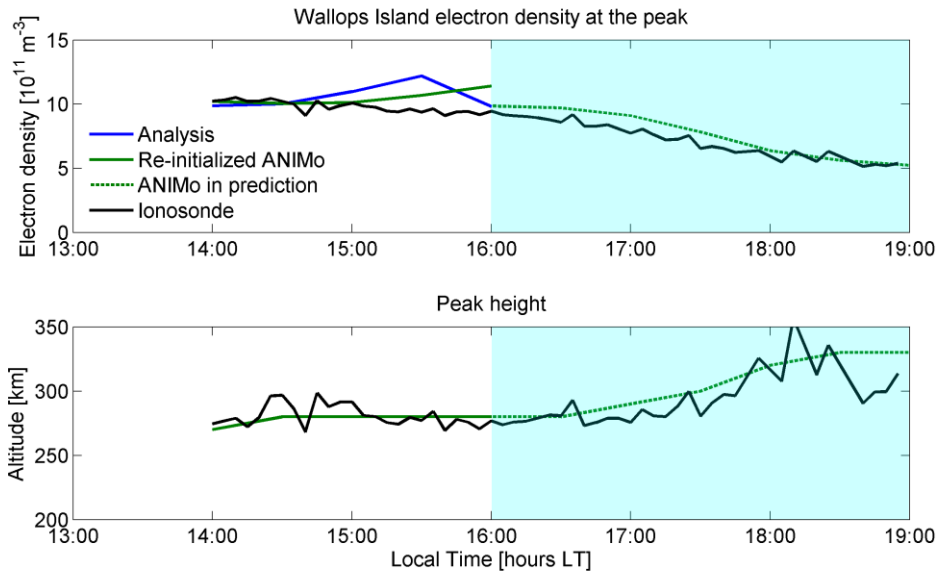

Figure 38. The graphs show, respectively, the comparisons of electron densities at the peak and peak heights produced by ANIDAS method A (blue), by re-initialized ANIMo (green) and measured by ionosonde (black) for the now-casting case study (29th of December 2011 from 1400 to 1600 LT (1900 to 2100 UT)) and by ANIMo (dashed green) in forecasting (29th of December 2011 from 1600 to 1900 LT (2100 to 0000 UT), highlighted in light blue) over Wallops Island.

The dashed green lines in Figure 36, Figure 37, and Figure 38 refer to the forecasting results provided by ANIMo. In each location the model in prediction is behaving well in mimicking the ionosonde measurments and their trend in time.

8.3.1 Boulder case

The situation over Boulder is of particular interest. This is where the satellite observations are abundant. Figure 39 shows a further extension in time of the graph of Figure 36 with the addition of the outcomes from IRI-2012 (solid red line) focussing on the electron density at the peak. As for the previuos images, the 'future' is emphatized by a light blue color and the result of ANIMo prediction is outlined as a dashed green line.

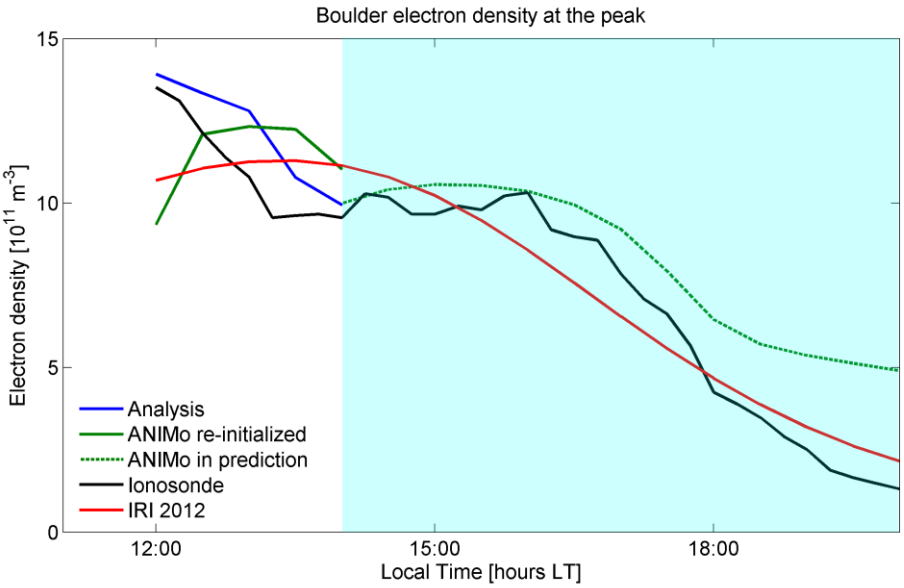

Figure 39. Comparisons of electron densities at the peak by ANIDAS method A (blue), by re-initialized ANIMo (green), by IRI-2012 (red) and measured by ionosonde (black) for the now-casting case study (29th of December 2011 from 1200 to 1400 LT (1900 to 2100 UT)) and by ANIMo (dashed green) in forecasting (29th of December 2011 from 1400 to 2000 LT (2100 to 0300 UT), highlighted in light blue) over Boulder.

The comparison of NmF2 values between IRI-2012 and the ionosonde clearly confirms what was anticipated in Section 8.2.1: here the ionosphere is behaving in an unexpected manner. This is true especially in the initial hours of the selected time window, but a certain mismatch is persistent also in the dusk period when the two NmF2 values seems to fade with different trends. In this situation the results of ANIDAS in now-casting and forecasting mode appear to be more accurate. In the final part of the time window (after 1700 LT), ANIMo prediction increasingly loses accuracy while IRI-2012 seems to recover it. This is confirmed by comparing the absolute differences between ANIDAS and IRI-2012 results against the ionosonde mesurement(Figure 40).

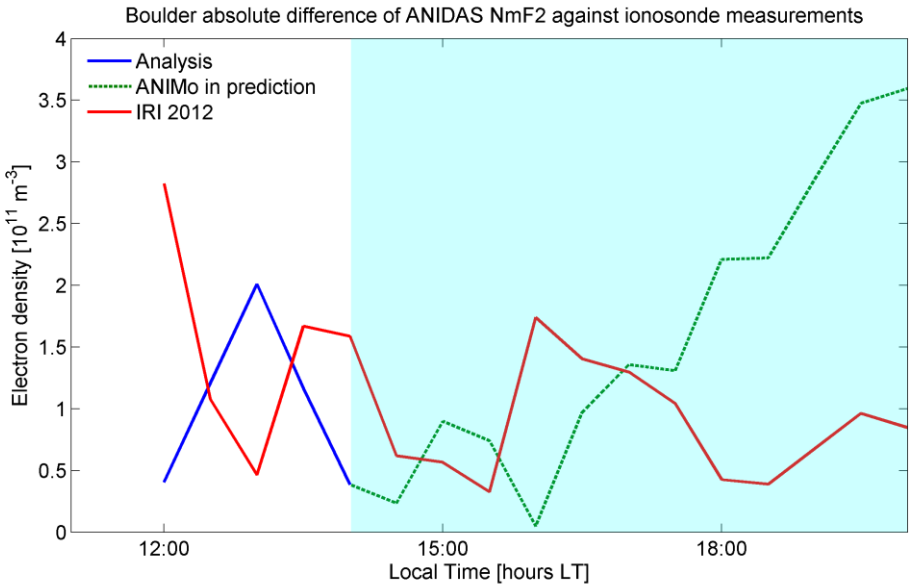

Figure 40. Comparisons of the absolute difference in electron density at the peak between ANIDAS method A (blue) and IRI-2012(red) against the ionosonde measurement for the now-casting case study (29th of December 2011 from 1200 to 1400 LT (1900 to 2100 UT)), and between ANIMo in forecasting (dashed green) and IRI-2012 (red) against the ionosonde (black) for the forecasting period (29th of December 2011 from 1400 to 2000 LT (2100 to 0300 UT), highlighted in light blue) over Boulder.

Figure 40 shows the mentioned differences; the blue and dashed green line refer to ANIDAS in now-casting and forecasting respectively, the red one is associated with IRI-2012. In this type of graph the lower the value the more accurate is the reconstruction. In the first two hours the blue line of ANIDAS now-casting runs beneath the red line of IRI-2012 for most of the time steps. This pattern is also found in between 1400 LT and 1700 LT when ANIDAS is used in prediction. During the last three hours of the time window, the difference between the forecasting and the ionosonde measurement increases. This offset was expected since the benefit of the initialization on a model accuracy is generally believed to diminish with time.

Summary

The preliminary results reported in this chapter show great potential. They demonstrated that, for the selected case study, all the goals of the project were achieved.

The general objective of the present project is to use a physics-based model in support ionospheric tomography imaging through a DA scheme. In particular, the setup has to provide new solutions for two ionospheric tomography limitations: the lack of data issue and the poor vertical resolution. According to the preliminary results displayed in this chapter, ANIDAS is performing resonably well in situations where GPS satellite observations are abundant and can also deal with different grades of lack of data at the same time. The reconstruction of the vertical electron density profiles and the good agreement of the electron density at the peak and peak altitude with ionosonde and ISR measurements reveal that ANIDAS can provide more accurate vertical specifications of the

ionosphere. These aspects are proved to be true also for consecutive time steps and, moreover, it is demostrated that there is a real benefit in the re-initialization of ANIMo in producing the new background state at each time. This is not only advantageous for the now-casting mode of ANIDAS, if ANIMo is properly initialized it can perform a good forecasting for several hours. This is also very relevant as one of the secondary objectives of this project was to use the DA scheme in oder to provide predictions of the ionosphere.

The validation of the presented scheme needs to be extended to additional case studies. In particular, it is expected to evaluate ANIDAS perfomances in different mid-latitudes regions, in different periods of the year and finally in different solar activity and geomagnetic conditions. Further investigations are required to better undestand the source (or sources) of the edge effects strongly visible in the TEC maps (Figure 27) of the 'cold start', which gradually fade in the consecutive hours (Figure 31). As aforementioned, one of the possible causes is the absence of a realistic boundary conditions in the background covariance matrix. This is a proposed future modification. Another critical aspect is the definition of the α parameter for each now-casting reconstruction. Future development should include a way to estimate an optimazed average α parameter.

Whether ANIDAS will exhibit good behaviour during the specified validation tests and will be modified to include the above improvements, it can be promoted to be used in an operative manner.

Chapter 9 Conclusions and future work

The design, development and implementation of a new physics-based ionospheric model for tomographic imaging and forecasting purposes were the principal aims of this PhD project.

In order to realise these aims a series of intermediate objectives needed to be achieved.

The first intermediate objective consisted of developing a new ionospheric model. The key points were to develop a model that contained the basic level of physics to produce a realistic vertical profile of the Earth's ionospheric electron density at mid-latitudes under quiet geomagnetic conditions. Robustness and reliability were also very important, because the model needed to be run automatically and routinely. Finally, the efficiency in terms of computational effort was essential. The second intermediate objective was to test the model against the above requirements. This involved the comparison of the model outputs against measurements of the ionosphere coming from instruments and other established models. The third intermediate objective addressed the implementation of the new model by means of a DA scheme.

ANIMo [*Da Dalt et al.*, 2014], described in Chapter 5 can be considered as the major achievement of the first intermediate objective. ANIMo is a physics-based model that produces three-dimensional reconstructions of the Earth's ionosphere in terms of electron density and the density of the major ions (O^+, O_2^+, and NO^+). Its outputs apply to anywhere on the Earth however the encoded physics are designed for mid-latitude regions. It provides outputs within an a range from 80 to 600 km altitude.

The continuity equation of O^+ is the core of ANIMo, it considers production given by photoionization, loss from chemical recombination and the transport processes due to an adjusted ambipolar diffusion. The determination of the time and location of the desired reconstruction, plus *F10.7* and A_p indexes, boundary conditions and the vertical velocity adjustment represent the list of inputs for ANIMo. The indexes are automatically retrieved and the boundary conditions and velocity adjustments are responsible for the refinements of the final result, thus ANIMo can be considered a user-friendly model. The small amount of physical processes required also makes ANIMo an extremely useful tool for experimenting with new ionospheric modelling solutions.

The accuracy of ANIMo was tested through comparisons (Section 6.2) with observations of electron density at the peak and peak altitude from both ionosonde and ISR measurements, and simulations by IRI-2012. The tests were performed above the location of the Millstone Hill Haystack observatory for multiple consecutive days in each seasons of the year under quiet

conditions. Chapter 6 reports the results of these tests and, in general ANIMo reproduces NmF2 and hmF2 well following diurnal and seasonal changes. In particular, the RMS difference of NmF2 estimated by ANIMo and IRI-2012 against the NmF2 measured by the ionosonde was comparable: by averaging the NmF2 value for all the unperturbed cases, IRI-2012 was more accurate by just 0.109 10^{11} electrons m^{-3}. By performing the same operation with regards to hmF2, IRI-2012 was more accurate than ANIMo by 9.472 km on average. ANIMo's accuracy was also tested in unsettled geomagnetic conditions. The results displayed in Subsection 6.2.1 demonstrate that, although the model did not accurately simulate the perturbed ionospheric features, it was able to provide a plausible reconstruction.

It has to be said that the top-side boundary condition was set to reach this level of accuracy. However, considering that IRI-2012 is adopted as standard reference for the ionosphere, taking into account that ANIMo is a first-principle model based on a small amount of physical processes, and that its vertical resolution cannot be lower than 10 km, the results for ANIMo's validation in this particular case study are very encouraging. This was confirmed by a personal communication with J. J. Sojka in the final stages of ANIMo development (October 2013). On this occasion, the winter and summer cases were compared to outcomes given by the USU TDIM [*Sojka et al.*, 2013].

Temperature sensitivity testing (Section 6.3) was performed in order to assess ANIMo's robustness in the sense of reliability in response to extreme driving forces inputs. These tests proved that ANIMo can deal reliably with extreme values of electron and ion temperature. This feature of ANIMo is crucial when using the model for driving forces experimentations. Another interesting result, for the winter case, was the very small difference between using IRI-2012 simulated temperatures, using the measured ones by ISR and setting to a constant 1000 K as input. Sensitivity tests were also performed for other input parameters: the top-side boundary condition (Section 6.4) and the vertical velocity adjustment (Section 6.5). The results showed how ANIMo was affected by different input settings and that a better representation of these parameters could improve the model's accuracy. The efficiency of ANIMo was also evaluated and improved on in various stages of its development, however no test results were reported because not directly pertinent to the topics treated in this dissertation. The validation and, the positive assessment of the ANIMo's required features represent the successful achievements for the second intermediate objective.

The realization of ANIDAS and of the two methods (A and B) for generating vertical basis functions, described in Chapter 7, successfully fulfilled the third intermediate objective. ANIDAS was also applied and tested, as explained in Chapter 8. The chosen location was the US area because of the ionospheric instrumentation at mid-latitudes. The selected case study was that of winter 2011, previously used in ANIMo validation tests. This case was chosen because of the good quality and availability of measurements – data from the ionosondes in Boulder, Wallops Island

and Millstone Hill were available (for the last location also ISR data). Furthermore the conditions presented in this case study, such as medium solar intensity and unperturbed geomagnetic activity, proved to be most suitable for an initial validation test. From ANIMo's validation tests, a stable regular behaviour of the ionosphere with a balanced diurnal trend and substantial electron density content were also observed. Furthermore, the GPS ground-based receivers' distribution was very good; however data from the Eastern sector were intentionally omitted for this test. The plan was to create a situation of simultaneous different degrees of data coverage to test the performances of ANIDAS. The winter case was the only one considered because of the computational load required in performing the presented validation test.

The first analysis, referred to as 'cold start', was set at 1900 UT on the 29th of December 2011. As testing criteria, the resulting TEC maps and the comparisons of the density profiles, and their characteristic parameters NmF2 and hmF2, were compared with other models and instruments. The 'cold start' TEC map (Figure 27) revealed that ANIDAS was capable of combining ANIMo background well with the observation contribution. However the reconstruction was quite noisy at the edges. By looking at the profiles, it was possible to notice that ANIDAS was very accurate and that the usage of methods A and B did not produced any noticeable difference for this case. The accuracy was confirmed by the NmF2 and hmF2 comparisons. ANIDAS and IRI results were comparable. For Boulder location where the data coverage was good, ANIDAS performed much better in terms of NmF2. Regarding hmF2, IRI-2012 seemed for the three locations to be slightly more accurate (on average ANIDAS differs from IRI-2012 by 4.867 km with respect to the ionosonde).

The now-casting capability of ANIDAS was also tested by advancing the DA reconstructions by 30 minutes time steps for two hours and comparing NmF2 and hmF2 results with all three ionosondes data (Subsection 8.3.2). Here the background model ANIMo used the analysis of the previous step as initialization, method A was used and the data on the Eastern sector were still omitted. The results demonstrate that ANIDAS performs well for both NmF2 and hmF2 parameters in all the locations. To be precise, in Millstone Hill ANIDAS overestimated the peak altitude by a consistent offset. The TEC maps at each time step showed that the noisy border effect observed at the 'cold start' gradually faded as the scheme advances. The evolution of the model from the previous analysis increased the accuracy and smoothness of the analysis at the current time. The reconstruction was hence less distressing than in the 'cold start' when a significant disagreement between observations and background was present (Figure 25 and Figure 41).

A very important aspect was the vertical constraint approach that was applied throughout the ANIDAS now-casting validation test. Method A was chosen for the reasons explained in Section 7.3.1. This approach uses a priori information extrapolated from ANIMo background to aid the

inversion in reconstructing the vertical profile. Its function is very relevant, as the ionospheric measurements from GPS ground-based receivers cannot provide a good vertical resolution. Where the data coverage was very low and there were very few observations, Method A simply tended to match the background contribution.

During ANIDAS validation the α parameter was manually adjusted for obtaining plausible reconstructions. Further work is planned to make this operation automatic.

ANIDAS can also be used in forecasting modality (Section 8.3). The last analysis was used to initialise ANIMo to run for future time steps, each one 30 minutes long. The forecasted TEC maps showed a predicted behaviour of the ionosphere; i.e. a gradient was present and consistent through all the prediction time and with the now-casting TEC maps. The veracity of the prediction was once again tested against the available ionosonde data. In all the locations, ANIMo was capable of providing a reliable forecast that faithfully followed the measurements. The peak altitude provided by ANIMo for Millstone Hill location appeared overestimated and the offset was consistent with the one observed in the now-casting validation test. Particularly interesting is the situation over Boulder (Subsection 8.3.1). Here the disagreement between the NmF2 from the ionosonde and extrapolated from IRI-2012 suggests an unusual behaviour of the ionosphere. ANIDAS coped with this in now-casting and ANIMo maintained a definite accuracy in forecasting it. After approximatively three hours the prediction starts to worsen by deviating from the ionosonde data. The loss in accuracy was expected as the advantage of the initialization is known to have a temporary effect. These comparisons represented the achievement of the last project intermediate objective and, most importantly, demonstrated that ANIDAS now-casting and forecasting modalities provided a reliable characterization and prediction of the ionosphere for the chosen case study.

The presented project results are very promising; however there are a series of potential tasks and further developments that need to be considered.

Regarding ANIMo, an extension of its validation test is needed to verify the accuracy of the model in other mid-latitude locations and during more perturbed space weather conditions. The simulation of the chemical processes that govern the bottom of the vertical profile can be improved. Figure 28, Figure 29, and Figure 30 clearly showed that ANIMo, compared to other methods, was not capable of reproducing the structures of the lower part of the electron density profile. Finally, it might be interesting to introduce a self-consistent calculation of the temperature in place of the contribution from IRI-2012 in order to make ANIMo less dependent on empirical information. Further improvements, such as the introduction of ionospheric wind in the transportation processes, the extension to polar and equatorial regions, the coupling with neutral atmosphere and magnetosphere dynamics should also be considered.

Further validation tests are also necessary for ANIDAS. In this project, solely one case study was selected because of the high computational effort required. For a more rigorous validation, the test should be extended to other periods of the year, other locations and during perturbed conditions of the ionosphere. At the current stage the generator of the background covariance matrix could be improved. It is based on the correlation distances statistics from mainly mid-latitude ionosondes collected by *Rush* [1976] and ordered by LMT and time of the year. A major step forward would be to include more correlation distances studies from different instruments (e.i. CHAMP satellite) and imaging techniques in order to take into account different solar and geomagnetic conditions and to go over the current vertical approximation. A proper boundary condition could also be included and tested to determine whether it has any influence on the border effect present on the 'cold start' case. Regarding this effect, further investigations are needed. A plausible solution however, would be to start the assimilation before the given time and discard the first reconstructions.

There are various exciting developments for this project at hand. Notwithstanding, this dissertation showed that the approaches presented produce plausible, reliable and stable results. The concept now is to continue using these techniques as workbenches to develop new solutions which would be helpful for the whole ionospheric now-casting and forecasting scientific community.

Appendix A

A.1 MIDAS, the calibration dilemma and the role of Lambda parameter

The calibration issue raised another interesting point of discussion: given that MIDAS can work with residual values, would the final result and measurement offsets be the same compared to the one performed by the canonical MIDAS version? According to the demonstrations reported on Section 7.1.1, the residual version of MIDAS could be written in the following formulation:

$$J_{M\text{-}r}(\delta \boldsymbol{x_r}) = \|\delta \boldsymbol{z} - \mathbf{H}\mathbf{K}\delta \boldsymbol{x_r}\|^2 + \lambda \left\|\boldsymbol{\nabla^2}\mathbf{K}\delta \boldsymbol{x_r}\right\|^2 \tag{A.1}$$

Similarly to the demonstration from Section 7.1.1, it is possible to derive a non-residual version of MIDAS from the residual described by Equation (A.1):

$$J_{M\text{-}nr}(\boldsymbol{x_{nr}}) = \|\boldsymbol{z} - \mathbf{H}\mathbf{K}\boldsymbol{x_{nr}}\|^2 + \lambda \left\|\boldsymbol{\nabla^2}\mathbf{K}\boldsymbol{x_{nr}} - \boldsymbol{\nabla^2}\boldsymbol{n}_b\right\|^2 \tag{A.2}$$

By comparing the 'non-residual' version of MIDAS (Eq. (A.2)) to the canonical version (Eq. (7.8)), a difference is noticed in the regularization term. The non-residual version exhibits in effect the presence of the background term $\boldsymbol{n_b}$. The regularization is performed here by taking into account an absolute value provided by the background rather than rely on relative values, as in the canonical version. This means that if the background has an unexpected strong gradient, the regularization will have to regard it. Therefore it is possible to conclude that MIDAS can work with residual values but the final result and the offsets could be different from the canonical version. An experiment was setup to establish whether this represents an issue for the implementation of ANIDAS. The chosen case of study is the same as the one used for the 'cold case', Section 8.2.1, and this includes the simulated low data coverage. MIDAS was used to reconstruct the ionosphere over US with the support of IRI model for producing the vertical profile basis functions (2 EOFs were used). Two different arrangements were followed: a) standard MIDAS, b) standard MIDAS applied to observation residual values produced as shown in Equation (A.1) from the realization of ANIMo. The following pictures (Figure 41) show the comparison between the TEC maps resulting from the two arrangements. Three couples of reconstructions were produced by changing the λ parameter which weights the contribution of the regularization factor within the inversion.

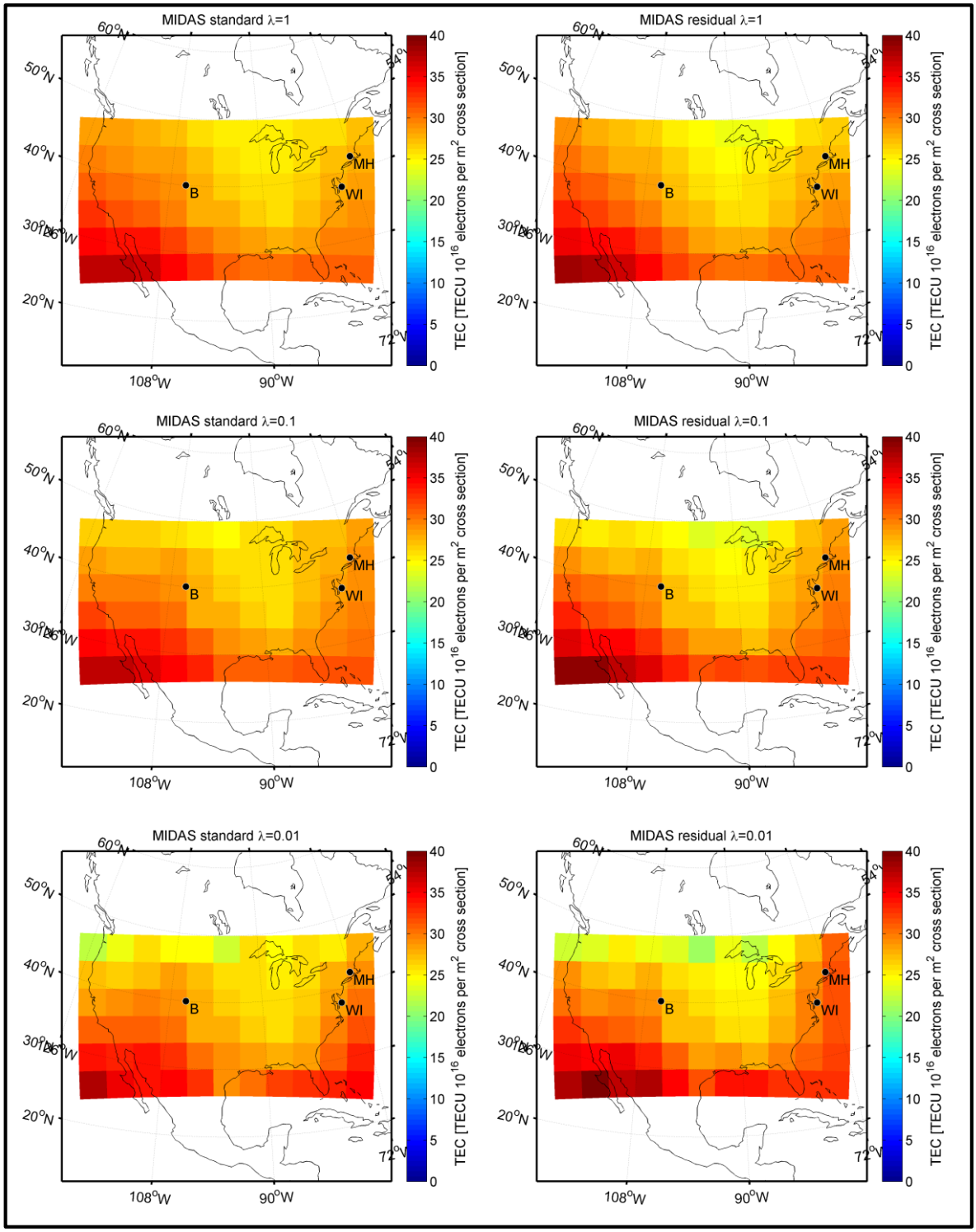

Figure 41. TEC maps from the reconstructions calculated by the standard version of MIDAS (on the left) and the residual version of MIDAS (on the right) by using decreasing values of Lamba (from top to the bottom) for the 'cold start' case study (29th of December 2011 at 1900 UT)

From Equation (7.7) in Section 7.1, the modelled observation $\boldsymbol{z_b}$ are considered unbiased because they are inferred from the model which is assumed to be exempt of biases. The offsets associated with the observation $\boldsymbol{z}$ is therefore carried within the inversion algorithm by the residual value $\boldsymbol{\delta z}$. The three couples of TEC maps displayed in Figure 41 do not exhibit significant differences

between themselves. This collateral experiment demonstrated that the MIDAS calibration method is able to deal with the observation offsets even if they are associated to a residual value. Nevertheless, there are small dissimilarities between the MIDAS standard and the residual reconstruction in agreement with the demonstration proposed in this section.

Figure 41 is also useful to understand the rule of the λ parameter and in general of the regularization factor in MIDAS, which was explained in Section 7.1. As often happens, the definition of λ is in part arbitrary but it should be related with the condition of the ionosphere at the moment of the reconstruction and on the data coverage. The TEC maps shown in Figure 41 refers to a relatively calm situation where a smooth ionosphere where taking place. This is confirmed by the fact that the change of λ, and therefore the tuning of the weight of the contribution of the regularization, makes very little difference in the results. Obviously, this is not necessarily true in other conditions.

Literature cited

AIAA (1999), Guide to Reference and Standard Ionosphere Models *Rep.*, American Institute of Aeronautics and Astronautics, Reston, Virginia, US.

Anderson, D. N. (1973), A theoretical study of the ionospheric *F* region equatorial anomaly—I. Theory, *Planetary and space science*, *21*(3), 409-419.

Anderson, D. N., M. Mendillo, and B. Herniter (1987), A semi-empirical low-latitude ionospheric model *Radio Science*, *22*(2), 292–306.

Anderson, D. N., et al. (1998), Intercomparison of physical models and observations of the ionosphere, *Journal of Geophysical Research: Space Physics (1978–2012)*, *103*(A2), 2179-2192.

Angling, M. J., and P. S. Cannon (2004), Assimilation of radio occultation measurements into background ionospheric models, *Radio Science*, *39*(1), RS1S08.

Angling, M. J., and B. Khattatov (2006), Comparative study of two assimilative models of the ionosphere, *Radio science*, *41*(5), RS5S20.

Bailey, G., and R. Sellek (1990), A mathematical model of the earth's plasmasphere and its application in a study of He (+) at L= 3, *Annales Geophysicae*, *8*(3), 171-189.

Bailey, G., R. Sellek, and Y. Rippeth (1993), A modelling study of the equatorial topside ionosphere, *Annales geophysicae*, *11*(4), 263-272.

Bailey, G., N. Balan, and Y. Su (1997), The Sheffield University plasmasphere ionosphere model--a review, *Journal of Atmospheric and Solar-Terrestrial Physics*, *59*(13), 1541-1552.

Barclay, L. W. (2003), *Propagation of radiowaves*, Institution of Electrical Engineers, London, UK.

Benton, C. J., and C. N. Mitchell (2012), New method for tracking the movement of ionospheric plasma, *Journal of Geophysical Research: Space Physics (1978–2012)*, *117*(A9), A09317.

Beynon, W., and P. Williams (1978), Incoherent scatter of radio waves from the ionosphere, *Reports on Progress in Physics*, *41*(6), 909-955.

Bilitza, D. (1990), International Reference Ionosphere 1990 *Rep.*, NSSDC 90-22, World Data Center A, Rockets and Satellites, Greenbelt, Mariland, US.

Bouttier, F., and P. Courtier (2002), Data assimilation concepts and methods March 1999, *Meteorological training course lecture series,* European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK.

Bowles, K. (1958), Observation of vertical-incidence scatter from the ionosphere at 41 Mc/sec, *Physical Review Letters*, *1*(12), 454.

Breit, G., and M. A. Tuve (1925), A radio method of estimating the height of the conducting layer, *Nature*, *116*, 357.

Bust, G. S., T. W. Garner, and T. L. Gaussiran II (2004), Ionospheric Data Assimilation Three-Dimensional (IDA3D): A global, multisensor, electron density specification algorithm, *Journal of geophysical research*, *109*(A11), A11312.

Bust, G. S., G. Crowley, N. Curtis, A. Reynolds, L. Paxton, C. Coker, and P. Bernhardt (2007), IDA4D-a new ionospheric imaging algorithm using non-linear ground-based and spaced-based data sources, paper presented at AGU Fall Meeting Abstracts.

Da Dalt, F., C. J. Benton, T. Panicciari, N. D. Smith, and C. N. Mitchell (2014), ANIMo — A New Ionospheric Model. Ionospheric Modelling for Ionospheric Imaging and Forecasting Purposes, in *Mitigation of Ionospheric Threats to GNSS: an Appraisal of the Scientific and Technological Outputs of the TRANSMIT Project*, edited by R. Notarpietro, pp. 201-209, InTech.

Daniell Jr, R. E. (1991), Parameterized Real-Time Ionospheric Specification Model PRISM Version 1. 0 *Rep.*, Phillips Lab., Hanscom Air Force Base, Massachussetts, US.

Daniell Jr, R. E., and L. D. Brown (1995), PRISM: A Parameterized Real-Time Ionospheric Specification Model, Version 1.5 *Rep.*, Phillips Lab., Hanscom Air Force Base, Massachussetts, US.

Daniell Jr, R. E., L. Brown, D. Anderson, M. Fox, P. Doherty, D. Decker, J. J. Sojka, and R. W. Schunk (1995), Parameterized ionospheric model: A global ionospheric parameterization based on first principles models, *Radio Science*, *30*(5), 1499-1510.

Davies, K. (1990), *Ionospheric radio*, Peter Peregrinus Ltd, London, UK.

Decker, D., C. Valladares, R. Sheehan, S. Basu, D. Anderson, and R. Heelis (1994), Modeling daytime F layer patches over Sondrestrom, *Radio Science*, *29*(1), 249-268.

Di Giovanni, G., and S. M. Radicella (1990), An analytical model of the electron density profile in the ionosphere, *Advances in Space Research*, *10*(11), 27-30.

Dickinson, R. E., E. Ridley, and R. Roble (1981), A three-dimensional general circulation model of the thermosphere, *Journal of Geophysical Research*, *86*(A3), 1499-1512.

Drob, D. P., J. T. Emmert, G. Crowley, J. M. Picone, G. G. Shepherd, W. Skinner, P. Hays, R. J. Niciejewski, M. Larsen, and C. Y. She (2008), An empirical model of the Earth's horizontal wind fields: HWM07, *Journal of Geophysical Research: Space Physics (1978–2012)*, *113*(A12), A12304.

Fennelly, J. A., and D. G. Torr (1992), Photoionization and photoabsorption cross sections of O, N2, O2, and N for aeronomic calculations, *Atomic Data and Nuclear data tables*, *51*(2), 321-363.

Finlay, C. C., S. Maus, C. D. Beggan, T. N. Bondar, A. Chambodut, T. A. Chernova, A. Chulliat, V. P. Golovkov, B. Hamilton, and M. Hamoudi (2010), International geomagnetic reference field: the eleventh generation, *Geophysical Journal International*, *183*(3), 1216-1230.

Fougere, P. F. (1995), Ionospheric radio tomography using maximum entropy 1. Theory and simulation studies, *Radio science*, *30*(2), 429-444.

Fremouw, E., J. A. Secan, and B. M. Howe (1992), Application of stochastic inverse theory to ionospheric tomography, *Radio Science*, *27*(5), 721-732.

Fuller-Rowell, T. J., and D. Rees (1980), A three-dimensional time-dependent global model of the thermosphere, *Journal of the Atmospheric Sciences*, *37*(11), 2545-2567.

Fuller-Rowell, T. J., D. Rees, S. Quegan, R. Moffett, M. Codrescu, and G. Millward (1996), A coupled thermosphere-ionosphere model (CTIM), in *STEP: Handbook of Ionospheric Models*, edited by R.W. Schunk, pp. 239-279, Utah State University, US.

Gordon, W. E. (1958), Incoherent scattering of radio waves by free electrons with applications to space exploration by radar, *Proceedings of the IRE*, *46*(11), 1824-1829.

Hajj, G. A., B. D. Wilson, C. Wang, X. Pi, and I. G. Rosen (2004), Data assimilation of ground GPS total electron content into a physics-based ionospheric model by use of the Kalman filter, *Radio Science*, *39*(1), RS1S05.

Hargreaves, J. (1992), *The Solar-Terrestrial Environment*, Cambridge University Press, New York, US.

Harris, M. J. (2000), A new coupled terrestrial mesosphere-thermosphere general circulation model: Studies of dynamic, energetic, and photochemical coupling in the middle and upper atmosphere, PhD Thesis, University of London, UK.

Harris, M. J., N. F. Arnold, and A. D. Aylward (2002), A study into the effect of the diurnal tide on the structure of the background mesosphere and thermosphere using the new coupled middle atmosphere and thermosphere (CMAT) general circulation model, *Annales Geophysicae*, *20*, 225-235.

Hedin, A. E. (1987), MSIS-86 thermospheric model, *Journal of Geophysical Research*, *92*(A5), 4649-4662.

Hedin, A. E., N. W. Spencer, and T. L. Killeen (1988), Empirical global model of upper thermosphere winds based on Atmosphere and Dynamics Explorer satellite data, *Journal of Geophysical Research: Space Physics (1978–2012)*, *93*(A9), 9959-9978.

Huba, J., G. Joyce, and J. Fedder (2000), Sami2 is Another Model of the Ionosphere (SAMI2): A new low-latitude ionosphere model, *Journal of Geophysical Research*, *105*(A10), 23035-23053.

Jakowski, N. (1996), TEC monitoring by using satellite positioning systems, in *Modern ionospheric science*, edited by R. R. K. Kohl, and K. Schlegel, pp. 371-390, European Geophysical Society, Katlenburg-Lindau, Germany.

Jin, H., Y. Miyoshi, H. Fujiwara, H. Shinagawa, K. Terada, N. Terada, M. Ishii, Y. Otsuka, and A. Saito (2011), Vertical connection from the tropospheric activities to the ionospheric longitudinal structure simulated by a new Earth's whole atmosphere-ionosphere coupled model, *Journal of Geophysical Research*, *116*(A1), A01316.

Kohl, H., and J. W. King (1967), Atmospheric winds between 100 and 700 km and their effects on the ionosphere, *Journal of Atmospheric and Terrestrial Physics*, *29*(9), 1045-1062.

Leitinger, R., M. L. Zhang, and S. M. Radicella (2005), An improved bottomside for the ionospheric electron density model NeQuick, *Annals of Geophysics*, *48*(3), 525-534.

Mandrake, L., B. D. Wilson, G. Hajj, C. Wang, X. Pi, and B. Iijima (2004), USC/JPL GAIM: A Real-Time Global Ionospheric Data Assimilation Model, paper presented at AGU Fall Meeting Abstracts.

Mannucci, A. J., B. Iijima, U. Lindqwister, X. Pi, L. Sparks, and B. Wilson (1999), GPS and ionosphere, in *Review of Radio Science 1996–1999*, edited by W. R. Stone, pp. 625-665, Oxford Univ. Press, New York, US.

Matsushita, S., and W. H. Campbell (1967), *Physics of Geomagnetic Phenomena, Volume 1*, Academic Press, New York, US.

McNamara, L. F. (1991), *The ionosphere: communications, surveillance, and direction finding*, Krieger Publishing Company, Malabar, Florida, US.

Millward, G., R. Moffett, S. Quegan, and T. Fuller-Rowell (1996), A coupled thermosphere-ionosphere-plasmasphere model (CTIP), *STEP: Handbook of Ionospheric Models*, 239-279.

Millward, G., I. Müller-Wodarg, A. Aylward, T. Fuller-Rowell, A. Richmond, and R. Moffett (2001), An investigation into the influence of tidal forcing on F region equatorial vertical ion drift using a global ionosphere-thermosphere model with coupled electrodynamics, *Journal of Geophysical Research. A. Space Physics*, *106*(A11), 24733–24744.

Mitchell, C. N., and P. S. J. Spencer (2003), A three-dimensional time-dependent algorithm for ionospheric imaging using GPS, *Annals of Geophysics*, *46*(4), 687-696.

Namgaladze, A. A., Y. N. Korenkov, V. V. Klimenko, I. V. Karpov, F. S. Bessarab, V. A. Surotkin, T. A. Glushchenko, and N. M. Naumova (1988), Global model of the thermosphere-ionosphere-protonosphere system, *Pure and applied geophysics*, *127*(2), 219-254.

Nava, B., P. Coïsson, and S. M. Radicella (2008), A new version of the NeQuick ionosphere electron density model, *Journal of Atmospheric and Solar-Terrestrial Physics*, *70*(15), 1856-1862.

Nisbet, J. (1971), On the construction and use of a simple ionospheric model, *Radio Science*, *6*(4), 437-464.

Nisbet, J., and R. Divany (1992), Penn state Mk III model (1985), *Planetary and space science*, *40*(4), 545-545.

Peymirat, C., and D. Fontaine (1994), Numerical simulation of magnetospheric convection including the effect of field-aligned currents and electron precipitation, *Journal of Geophysical Research*, *99*(A6), 11155-11176.

Peymirat, C., A. Richmond, B. Emery, and R. Roble (1998), A magnetosphere-thermosphere-ionosphere electrodynamics general circulation model, *Journal of Geophysical Research*, *103*(A8), 17467-17477.

Pi, X., C. Wang, G. A. Hajj, G. Rosen, B. D. Wilson, and G. J. Bailey (2003), Estimation of E× B drift using a global assimilative ionospheric model: An observation system simulation experiment, *Journal of Geophysical Research: Space Physics (1978–2012)*, *108*(A2), SIA 12-11-SIA 12-13.

Picone, J., A. Hedin, D. P. Drob, and A. Aikin (2002), NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues, *Journal of Geophysical Research*, *107*(A12), SIA 15-11–SIA 15-16.

Quegan, S., G. Bailey, R. Moffett, R. Heelis, T. Fuller-Rowell, D. Rees, and R. Spiro (1982), A theoretical study of the distribution of ionization in the high-latitude ionosphere and the plasmasphere: first results on the mid-latitude trough and the light-ion trough, *Journal of Atmospheric and Terrestrial Physics*, *44*(7), 619-640.

Rawer, K., D. Bilitza, and S. Ramakrishnan (1978), Goals and status of the International Reference Ionosphere, *Reviews of Geophysics*, *16*(2), 177-181.

Raymund, T., S. Franke, and K. Yeh (1994), Ionospheric tomography: its limitations and reconstruction methods, *Journal of atmospheric and terrestrial physics*, *56*(5), 637-657.

Rees, M. H. (1989), *Physics and chemistry of the upper atmosphere*, Cambridge University Press, New York, US.

Richards, P. G., J. A. Fennelly, and D. G. Torr (1994a), Correction to "EUVAC: A solar EUV flux model for aeronomic calculations", *Journal of Geophysical Research*, *99*(A7), 13283.

Richards, P. G., J. A. Fennelly, and D. G. Torr (1994b), EUVAC: A solar EUV flux model for aeronomic calculations, *Journal of Geophysical Research*, *99*(A5), 8981-8992.

Richards, P. G., D. G. Torr, M. E. Hagan, and M. J. Buonsanto (1995), A new algorithm for improved ionospheric electron density modeling, *Geophysical research letters*, *22*(11), 1385-1388.

Richmond, A., E. Ridley, and R. Roble (1992), A thermosphere/ionosphere general circulation model with coupled electrodynamics, *Geophysical research letters*, *19*(6), 601-604.

Ridley, A., and M. Liemohn (2002), A model-derived storm time asymmetric ring current driven electric field description, *Journal of Geophysical Research*, *107*(A8), SMP 2-1-SMP 2-12.

Ridley, A., Y. Deng, and G. Toth (2006), The global ionosphere–thermosphere model, *Journal of Atmospheric and Solar-Terrestrial Physics*, *68*(8), 839-864.

Rishbeth, H., and O. K. Garriott (1969), *Introduction to ionospheric physics*, Academic Press, New York, US.

Roble, R., and E. Ridley (1994), A thermosphere-ionosphere-mesosphere-electrodynamics general circulation model (time-GCM): Equinox solar cycle minimum simulations (30-500 km), *Geophysical research letters*, *21*(6), 417-420.

Roble, R., E. Ridley, A. Richmond, and R. Dickinson (1988), A coupled thermosphere/ionosphere general circulation model, *Geophysical research letters*, *15*(12), 1325-1328.

Rush, C. M. (1976), An ionospheric observation network for use in short-term propagation predictions, *ITU Telecommunication Journal*, *43*(8), 544-549.

Salah, J. E. (1993), Interim standard for the ion-neutral atomic oxygen collision frequency, *Geophysical research letters*, *20*(15), 1543-1546.

Salah, J. E., and J. M. Holt (1974), Midlatitude thermospheric winds from incoherent scatter radar and theory, *Radio Science*, *9*(2), 301-313.

Scherliess, L., R. W. Schunk, J. J. Sojka, and D. C. Thompson (2004), Development of a physics-based reduced state Kalman filter for the ionosphere, *Radio Science*, *39*(1), RS1S04.

Schunk, R. W. (1988), A mathematical model of the middle and high latitude ionosphere, *Pure and applied geophysics*, *127*(2), 255-303.

Schunk, R. W., and J. C. G. Walker (1973), Theoretical ion densities in the lower ionosphere, *Planetary and space science*, *21*(11), 1875-1896.

Schunk, R. W., J. J. Sojka, and J. V. Eccles (1997), Expanded capabilities for the ionospheric forecast model *Rep.*, AFRL-VS-HA-TR-98– 0001. Air Force Res. Lab., Hanscom Air Force Base, Massachussetts, US.

Schunk, R. W., L. Scherliess, J. J. Sojka, D. C. Thompson, and L. Zhu (2005a), An operational data assimilation model of the global ionosphere, paper presented at Ionospheric Effects Symposium Proceedings, Natl. Tech. Info. Serv., Springfield, Virginia, US.

Schunk, R. W., L. Scherliess, J. J. Sojka, D. C. Thompson, and L. Zhu (2005b), Ionospheric weather forecasting on the horizon, *Space Weather*, *3*(8), S08007.

Schunk, R. W., L. Scherliess, J. J. Sojka, D. C. Thompson, D. N. Anderson, M. Codrescu, C. Minter, T. J. Fuller-Rowell, R. A. Heelis, and M. Hairston (2004), Global assimilation of ionospheric measurements (GAIM), *Radio Science*, *39*(1), RS1S02.

Sojka, J. J. (1989), Global Scale, Physical Models of the F Region Ionosphere, *Reviews of Geophysics*, *27*(3), 371-403.

Sojka, J. J., M. David, and R. W. Schunk (2013), TDIM Comparison with Bath-Model for Summer and Winter 2011, Personal communication.

Sojka, J. J., W. Borer, D. N. Anderson, T. J. Fuller-Rowell, and R. W. Schunk (1995), The coupled ionospheric-thermospheric forecast model (CITFM): Backbone for real-time operational space weather forecasting. , paper presented at *NRL Workshop on Space Weather: Needs, Capabilities and Science*, Washington, DC.

Spencer, P. S. J., and C. N. Mitchell (2007), Imaging of fast moving electron-density structures in the polar cap, *Annals of Geophysics*, *50*(3), 427-434.

Sutton, E., and H. Na (1994), High resolution ionospheric tomography through orthogonal decomposition, paper presented at *IEEE International Conference on Image Processing*.

Tascione, T. F. (1994), *Introduction to the space environment*, Krieger Publishing Company, Malabar, Florida, US.

Tikhonov, A. N., V. I. Arsenin, and F. John (1977), *Solutions of ill-posed problems*, Winston Washington DC, US.

Torr, D. G., and M. R. Torr (1979), Chemistry of the thermosphere and ionosphere, *Journal of Atmospheric and Terrestrial Physics*, *41*(7-8), 797-839.

Torr, D. G., M. R. Torr, H. C. Brinton, L. H. Brace, N. W. Spencer, A. E. Hedin, W. B. Hanson, J. H. Hoffman, A. O. Nier, and J. C. G. Walker (1979), An Experimental and Theoretical Study of the Mean Diurnal Variation of O+, NO+, O2+, and N2+ Ions in the Mid-Latitude F 1 Layer of the Ionosphere, *Journal of Geophysical Research*, *84*(A7), 3360-3372.

Wang, C., G. Hajj, X. Pi, I. G. Rosen, and B. Wilson (2004), Development of the global assimilative ionospheric model, *Radio Science*, *39*(1), RS1S06.

Wang, W. (1998), A thermosphere-ionosphere nested grid (TING) model, PhD thesis, University of Michigan, US.

Wang, W., J. Lei, A. G. Burns, M. Wiltberger, A. D. Richmond, S. C. Solomon, T. L. Killeen, E. R. Talaat, and D. N. Anderson (2008), Ionospheric electric field variations during a geomagnetic storm simulated by a coupled magnetosphere ionosphere thermosphere (CMIT) model, *Geophysical research letters*, *35*(18), L18105.

Webb, P. A., and E. A. Essex (2001), A dynamic diffusive equilibrium model of the ion densities along plasmaspheric magnetic flux tubes, *Journal of Atmospheric and Solar-Terrestrial Physics*, *63*(11), 1249-1260.

Yeh, K., and T. Raymund (1991), Limitations of ionospheric imaging by tomography, *Radio Science*, *26*(6), 1361-1380.

Yigit, E., A. S. Medvedev, A. D. Aylward, P. Hartogh, and M. J. Harris (2009), Modeling the effects of gravity wave momentum deposition on the general circulation above the turbopause, *Journal of Geophysical Research*, *114*, D07101.